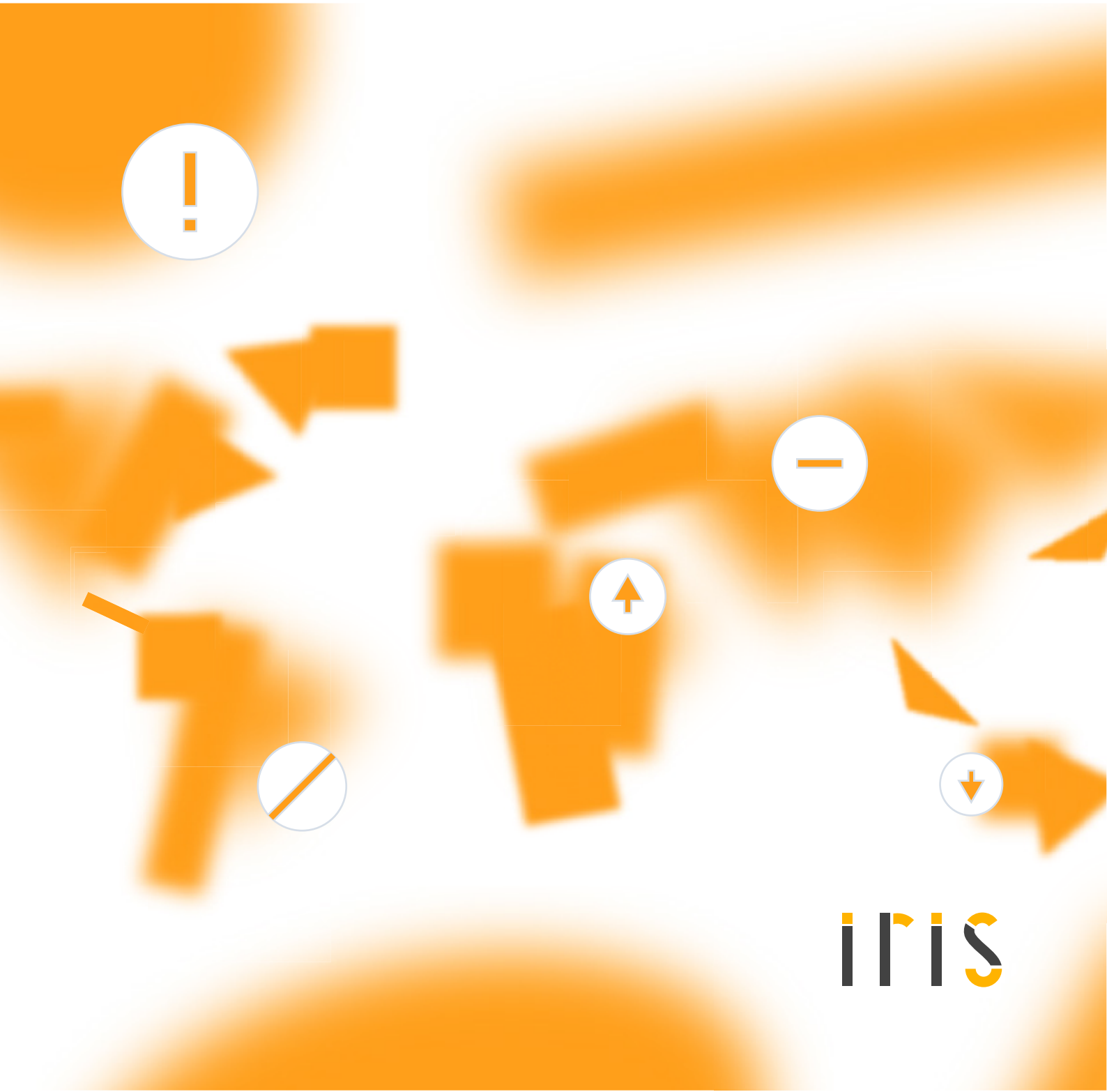
Transparência na

moderação de conteúdo

Tendências regulatórias nacionais



iris

Transparência na **moderação de conteúdo**

Tendências regulatórias nacionais

AUTORIA

Lahis Pasquali Kurtz

Paloma Rocillo Rolim do Carmo

Victor Barbieri Rodrigues Vieira

REVISÃO

Daphnee Prates Iglesias

João Archegas

Luíza Couto Chaves Brandão

PROJETO GRÁFICO, CAPA, DIAGRAMAÇÃO E FINALIZAÇÃO

Felipe Duarte

PRODUÇÃO EDITORIAL

Instituto de Referência em Internet e Sociedade

COMO CITAR EM ABNT

KURTZ, Lahis Pasquali; DO CARMO, Paloma Rocillo Rolim; VIEIRA, Victor Barbieri Rodrigues. **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: dd mmm. aaaa

A pesquisa contou com a revisão identificada de especialistas não integrantes da equipe do IRIS: Daphnee Prates Iglesias (experta em digitalização e políticas públicas; German Chancellor Fellow da Fundação Alexander von Humboldt) e João Archegas (Instituto Tecnologia e Sociedade), a quem agradecemos pelas valiosas contribuições

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Transparência na moderação de conteúdo : tendências regulatórias nacionais [livro eletrônico]. -- 1. ed. -- Belo Horizonte, MG : Instituto de Referência em Internet e Sociedade, 2021.
PDF

ISBN 978-65-87416-02-1

1. Ciências sociais 2. Internet (Rede de computador) 3. Plataforma digital 4. Regulação - Brasil 5. Transparência - Aspectos sociais.

21-71276 CDD-372.721

Índices para catálogo sistemático:

1. Regulação : Privacidade e proteção de dados pessoais : Direito 372.721

Aline Graziele Benitez - Bibliotecária - CRB-1/3129

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE

DIREÇÃO

Luíza Couto Chaves Brandão

MEMBROS

Ana Bárbara Gomes / Pesquisadora
Beatriz Fernandes / Estagiária de Comunicação
Felipe Duarte / Coordenador de Comunicação e Pesquisador
Gustavo Rodrigues / Coordenador de Políticas Públicas e Pesquisador
Juliana Roman / Pesquisadora
Lahis Kurtz / Coordenadora de Pesquisa e Pesquisadora
Leandro Soares Nunes / Pesquisador
Paloma Rocillo Rolim do Carmo / Diretora Financeira e Pesquisadora
Pedro Vilela Resende Gonçalves / Co-fundador e Associado
Victor Barbieri Rodrigues Vieira / Pesquisador

SUMÁRIO

RESUMO 6

INTRODUÇÃO 7

1. POR QUE MODERAÇÃO DE CONTEÚDO EXIGE TRANSPARÊNCIA? 9

1.1. Moderação de conteúdo 9

1.2. Principais práticas e técnicas utilizadas para moderação de conteúdo 12

a) Práticas de intervenção no conteúdo 14

b) Técnicas: Moderação algorítmica e humana 16

1.3. A transparência na moderação de conteúdo 18

1.3.1. O que é transparência? 18

1.3.2. Relação com o direito à liberdade de expressão 19

1.3.3. Importância da transparência algorítmica 23

1.3.4. Direito à privacidade e proteção de dados 25

1.4. Uma perspectiva crítica sobre a transparência 27

2. O PAPEL DA TRANSPARÊNCIA NO CENÁRIO REGULATÓRIO SOBRE CONTEÚDO ONLINE 30

2.1. Metodologia de análise 30

2.2. Aporte documental para requisitos de transparência 32

2.3. Apresentação dos resultados 36

2.3.1. Alemanha 36

a) Contexto político e midiático 36

b) Conteúdo da regulação 38

c) Requisitos de transparência 40

d) Obrigações quanto à moderação de conteúdo 41

2.3.2. Austrália 43

a) Contexto político e midiático 43

b) Conteúdo da regulação 45

c) Requisitos de transparência 47

d) Obrigações quanto à moderação de conteúdo 48

2.3.3. Brasil 48
a) Contexto político e midiático 48
b) Conteúdo da regulação 50
c) Requisitos de transparência 51
d) Obrigações quanto à moderação de conteúdo 55
2.3.4. Índia 55
a) Contexto político e midiático 55
b) Conteúdo da regulação 57
c) Requisitos de transparência 59
d) Obrigações quanto à moderação de conteúdo 61
2.3.5. Quênia 62
a) Contexto político e midiático 62
b) Conteúdo da regulação 65
c) Requisitos de transparência 66
d) Obrigações quanto à moderação de conteúdo 67

3. SÍNTESE COMPARATIVA 68
3.1. Tema e contexto das regulações 68
3.2. Papel da transparência na regulação analisada 70

4. CONCLUSÃO 77

REFERÊNCIAS 79

APÊNDICE A - REQUISITOS DE TRANSPARÊNCIA NA MODERAÇÃO DE CONTEÚDO 97

APÊNDICE B - QUADRO DE DISPOSITIVOS REGULATÓRIOS ANALISADOS 101

APÊNDICE C - FORMULÁRIO DE ANÁLISE REGULATÓRIA 103

Resumo

O objetivo desta pesquisa foi identificar padrões, lacunas e tendências nas obrigações de transparência impostas pela regulação estatal sobre moderação de conteúdo. Estudos anteriores demonstraram que as iniciativas mobilizadas voluntariamente pelas plataformas para publicizar as medidas adotadas carecem de robustez. Assim, o problema investigado é: as regulações aplicáveis a plataformas nos países de referência estipulam padrões obrigatórios de transparência sobre moderação de conteúdo? Para isso, a pesquisa teórica buscou identificar categorias relevantes para a moderação de conteúdo, bem como em que medida a transparência contribui para sua adequação a direitos fundamentais como livre expressão, acesso à informação e privacidade. Esta pesquisa foi complementada por estudo empírico, com coleta, análise e comparação da regulação sobre o tema em cinco países relevantes segundo o relatório Freedom on the Net: Alemanha, Austrália, Brasil, Índia e Quênia. Os parâmetros de análise basearam-se em referências internacionais, como Princípios de Manilla, Princípios de Santa Clara, relatórios como o Fostering Freedom Online e da Organização das Nações Unidas, entre outros. A regulação estudada denota significativa produção normativa aplicável a plataformas online em todos os países da amostra, especialmente voltada a controle de conteúdo danoso. As nomas analisadas, ao passo que amplificam o poder e os deveres dos intermediários da internet sobre o conteúdo que veiculam, não impõem medidas de transparência suficientes ao controle social sobre esta atividade. Isso dificulta a identificação das práticas adotadas, a discussão de seus reflexos e sugestão de aprimoramentos para o exercício de direitos na internet.

Palavras-chave: internet; moderação de conteúdo; plataformas; regulação; transparência

Introdução

A massificação da internet como fonte de informação e meio de expressão é também marcada por novos modelos de mídia. Um deles é a transformação comunicacional proporcionada por plataformas de conteúdo gerado por usuário, que traz dois elementos: 1. a ideia de uma mídia sem curadoria, ou seja, sem uma seleção inicial do que é disponível ao público ou quem são os comunicadores; 2. uma função de controle posterior que cresce cada vez mais em importância, à medida que as mesmas ferramentas que concedem a todas as pessoas a possibilidade de comunicar ideias e opiniões também potencializam os danos que alguns conteúdos podem causar. Os limites da autorregulação são postos à prova e emerge uma potencial área regulatória, voltada aos dilemas de controle de conteúdo, acesso à informação e livre expressão, que reconhece a transparência como importante fator de garantia desses direitos.

Frente a esse contexto, conhecer o tratamento da transparência pelas plataformas, por agentes da regulação e por usuários é fundamental na avaliação das medidas implementadas e necessidades emergentes. Nesta pesquisa, o enfoque é sobre o papel da regulação estatal no fomento à transparência nos serviços oferecidos por plataformas privadas na internet. Uma premissa é que a autorregulação[1] adotada como padrão inicial para esses serviços não se mostra suficiente. Já a regulação direta sobre o conteúdo pode ter efeitos colaterais à livre expressão. Uma solução balanceada seria uma regulação que garanta às instituições e aos usuários a fiscalização de medidas de controle de conteúdo online. Assim, o problema estudado é: os cenários regulatórios estatais impõem a adoção de práticas transparentes de moderação de conteúdo online que possibilitam sua fiscalização?

A fim de responder à questão, foram estudadas e comparadas as práticas regulatórias sobre transparência na moderação de conteúdo em plataformas digitais em diferentes jurisdições. Foram consideradas na análise as regulações nacionais e comunitárias[2], leis propostas ou já aprovadas no legislativo. A pesquisa abrangeu cinco jurisdições consideradas relevantes no contexto de cada um dos continentes – Alemanha, Austrália, Brasil, Índia e Quênia. Nesse processo, investigou-se quais as principais preocupações manifestadas em cada uma delas. Essa análise de material normativo (leis e projetos de leis) foi realizada à luz da fundamentação teórica que sustenta práticas institucionalizadas de transparência. Ainda, como parâmetros para criar categorias de comparação, foram identificadas as medidas recomendadas internacionalmente para a garantia de direitos humanos na moderação de conteúdo.

1 Ver mais em: BARTLE, Ian; VASS, Peter. **Self-regulation and the regulatory state: A survey of policy and practice**. Bath: Centre for the Study of Regulated Industries, University of Bath School of Management, 2005. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.3190&rep=rep1&type=pdf. Acesso em 15 jul. 2021.

2 Como as da União Europeia ou do MERCOSUL, que produzem regramentos a serem observados pelos estados-membros.

Esta pesquisa dá continuidade aos estudos sobre transparência na moderação de conteúdo realizados pelo IRIS, integrando o projeto "Transparência na moderação de conteúdo danoso: lei, usuários e plataformas"[3]. Neste estudo, buscou-se investigar um aspecto latente em análises de políticas de comunidade das plataformas: a necessidade de que as iniciativas de transparência sejam institucionalmente reforçadas. Isso porque não é suficiente a livre motivação dos atores privados em adotar os processos e medidas que julgam importantes. Ao identificar como a regulação aplicável aborda a transparência sobre moderação de conteúdo, podem-se reconhecer suas lacunas e sugerir recomendações para fortalecer um ambiente de confiança.

Percebe-se a emergência de uma preocupação estatal com o conteúdo online potencialmente danoso, em que afloram propostas regulatórias para sua moderação. Assim, é necessário averiguar se essas iniciativas também incluem, proporcionalmente ao poder de controle sobre o conteúdo conferido às plataformas, um poder fiscalizatório de controle da sociedade civil sobre como as plataformas implementam essas medidas. A regulação, assim, pode ser uma ferramenta institucional de garantia de adequação da moderação de conteúdo a parâmetros justos e democráticos, ou pode ampliar de forma desmedida o poder de agentes privados e públicos sobre a livre expressão. O estudo comparado da regulação pode auxiliar na identificação das tendências e diferenças contextuais nesse cenário.

Esta pesquisa está dividida em 3 partes, além desta introdução: a primeira seção discute as razões pelas quais há uma demanda por transparência sobre a atividade de moderação de conteúdo. Além disso, explica como ocorre essa atividade e seus desdobramentos práticos às pessoas usuárias de internet. Apresenta o conceito de transparência adotado, sua relação com direitos fundamentais e uma perspectiva crítica para as limitações de sua implementação na concretização desses direitos. A partir desse repertório, na segunda seção, delineia-se um estudo comparado de cinco diferentes regulações nacionais (Alemanha, Austrália, Brasil, Índia e Quênia) sobre moderação de conteúdo, por meio de pesquisa documental e formulário estruturado de análise da regulação localizada. Com base nesse material, são tecidas as conclusões.

3 IRIS - Instituto de Referência em Internet e Sociedade. **Transparência na moderação de conteúdo danoso:** lei, usuários e plataformas. Disponível em: https://irisbh.com.br/projetos/transparencia-na-moderacao-de-discurso-de-odio-lei-usuarios-e-plataformas/. Acesso em: 02 jan. 2020.

1. Por que moderação de conteúdo exige transparência?

Discussões sobre o papel das plataformas de conteúdo gerado por usuário e os limites que as interações podem ou devem ter estão cada vez mais na pauta do debate público. Com isso, percebe-se a importância de informação sobre o que é feito, e a necessidade de um diálogo embasado, que envolva relatos dos afetados pelas iniciativas de transparência sobre a moderação de conteúdo.

Nesta seção, apresentam-se as delimitações conceituais e a contextualização da moderação de conteúdo exercida por plataformas na internet. Também se discute o papel da transparência não só para melhoria dessas atividades como serviço, mas também como exercício de direitos.

1.1. Moderação de conteúdo

A intervenção sobre conteúdo é um tópico que permeia a agenda de questões da comunicação mediada. A moderação de conteúdo, entretanto, tem suas particularidades em relação ao tipo de intervenção realizada pela mídia tradicional, como televisão, rádio, jornal, entre diversas outras. A confecção de jornais impressos, por exemplo, envolve desde seus primórdios um trabalho de seleção das informações que serão veiculadas a cada edição: afinal, não é possível abordar todas as notícias existentes no período de publicação em um fino amontoado de papel. Adicionalmente, o trabalho editorial dos jornais tradicionais também envolve o posicionamento estratégico das informações, de forma a dar maior ou menor ênfase às mensagens veiculadas – o que por sua vez pode ser realizado para finalidades diversas.

Nesse contexto, a internet representou um meio no qual a intervenção sobre conteúdo alcançou novos patamares. Ao passo que a mídia tradicional realiza verdadeira curadoria, com uma demarcação bem delimitada entre quem produz conteúdo e quem é o público, nas redes sociais essa intervenção ocorre de outra forma. Um ponto de inflexão é a proposta das plataformas de conteúdo gerado por usuário, nas quais ocorre controle posterior sobre conteúdos específicos[4]. Isso pode se dar de maneira proativa (pelos moderadores da plataforma ou automaticamente pelos próprios algoritmos) ou motivada, seja por denúncias de participantes da comunidade[5],

4 Este é o aspecto do controle de conteúdo em plataformas que geralmente é denominado moderação de conteúdo. Ver: GILLESPIE, Tarleton L. et al. Expanding the Debate About Content Moderation.v. 9 n. 4**, Internet Policy Review**, 21 out. 2020

5 Também existem diferentes formatos, como a moderação realizada por usuários ou por funcionários da empresa ou terceirizados, de forma comercial. Ver: MYERS WEST, Sarah. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. **New Media & Society**, v. 20, n. 11, p. 4366-4383, 2018.

seja por ordens de autoridades locais, seja por pressão midiática. As plataformas de redes sociais se propõem como promotoras da livre expressão, tendo por ideal a possibilidade de qualquer pessoa circular qualquer conteúdo sem controle prévio. Entretanto, intervêm sobre o que é veiculado.

A moderação de conteúdo, portanto, é uma realidade para diversas plataformas que não selecionam previamente quem comunicará ou quais pautas serão divulgados em suas páginas. Esta atuação pode ser desempenhada de acordo com duas funções[6]. As plataformas podem ser "guardiãs" de conteúdo, pois determinam quais as categorias proibidas e permitidas; e podem ser organizadoras de conteúdo, pois ao destacar e priorizar alguns conteúdos em detrimentos de outros, realizam um trabalho semelhante ao de curadoria. Para tanto, as justificativas incluem a melhoria da experiência do usuário, a promoção da liberdade de expressão, a garantia de um modelo autorregulatório funcional e seguro, a luta contra ilícitos digitais, a viabilidade econômica da própria plataforma, entre outras[7]. Esses dois mecanismos – a moderação de conteúdo considerado inapropriado e a priorização de conteúdo considerado relevante – interferem na expressão propriamente dita, pois restringem ou ampliam o alcance de determinadas informações de maneira deliberada.

A moderação de conteúdo é um processo que envolve a negociação com uma pluralidade de agentes, além de quem administra e trabalha nas plataformas. Sander[8] afirma que ao menos quatro tipos de fatores impactam diretamente a moderação de conteúdo:

- Filosofia da empresa: cada plataforma desenvolve seus produtos para oferecer tipos específicos de experiência às pessoas que as usam. O Twitter, por exemplo, oferece a experiência de um microblog, enquanto que o Instagram possibilita o compartilhamento de fotos e vídeos. Portanto, os objetivos de cada plataforma permeiam os valores que serão praticados na moderação de conteúdo. A filosofia das plataformas serve tanto para criar uma base de pessoas usuárias quanto para satisfazer as missões pessoais daquelas que fundaram e que trabalham para a empresa.

6 SANDER, Barrie. Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation. **Fordham International Law Journal**, v. 43, p. 939, 2019. Disponível em: https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2787&context=ilj. Acesso em: 11 maio 2021

7 A autora Kate Klonick enumera três principais motivos que norteiam o modelo de moderação de conteúdo empregado nas principais plataformas digitais: 1. uma crença em normas de liberdade de expressão; 2. um senso de responsabilidade corporativa; e 3. uma necessidade de atender às normas dos usuários para viabilidade econômica. Ver: KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. **Harvard Law Review**, v. 131. n. 6. p. 1598-1670, abril de 2018. Disponível em: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Acesso em: 17 jun. 2021. p. 1625-1630.

8 Ibidem. p. 948-954.

- Conformidade regulatória: considerando que as plataformas são empresas submetidas à determinada jurisdição, mas atuantes em várias delas, diversas regulações devem ser observadas durante o processo de moderação. Além da pressão para cumprimento das normas vinculantes[9], existem pressões não-vinculantes que modelam a forma como as plataformas atuam. Este é o caso de uma unidade da Europol dedicada a encontrar conteúdo online potencialmente terrorista e enviá-lo para análise voluntária das plataformas[10]; ou ainda o Código de Conduta para Combate ao Discurso de Ódio firmado entre a Comissão Europeia e diversas plataformas[11].

- Maximização de lucros: A maior parte das plataformas contemporâneas estão baseadas na economia da atenção. Conforme argumenta Bentes[12], o modelo de negócio praticado pelas plataformas tem por base amplificar o engajamento na rede, o que possibilitará à plataforma coletar mais informações sobre o perfil de usuárias e usuários e oferecer-lhes serviços, produtos ou ideias que estão alinhados com os seus interesses individuais. Tal modelo de negócio está diretamente relacionado com o capitalismo de vigilância, teorizado por Zuboff[13], no qual as plataformas intermediam e facilitam a comunicação entre usuários em troca do monitoramento de suas expressões e comportamentos. É justamente a inserção de pessoas usuárias como matéria-prima dentro da cadeia de valor informacional que cria o excedente comportamental[14] (no texto original de Zuboff denominado *behavioral surplus*) que possibilita os lucros das plataformas. Por essa razão, muitas vezes, mudanças na moderação de conteúdo são consequências da demanda da indústria de anúncios. Em 2020, o boicote ao Facebook por

9 O termo "vinculante" refere-se a uma regra ou norma cuja inobservância possui consequências concretas para o agente implicado, sejam elas impostas por instituições do governo, da sociedade ou do mercado. Por exemplo, leis, normas e costumes de comércio internacional, diretrizes emanadas por instituições ou grupos que tenham atuação institucional. Não-vinculante é qualquer norma que não possui um reflexo direto decorrente da inobservância, mas que busca desencorajar determinadas situações ou condutas. Por exemplo, recomendações, princípios, padrões, códigos de conduta sem previsão de sanções.

10 EU INTERNET REFERRAL UNIT. **Transparency Report.** 2017. Disponível em: https://www.europol.europa.eu/publications-documents/eu-internet-referral-unit-transparency-report-2017. Acesso em: 14 jun. 2021

11 EUROPEAN UNION. **Code of conduct on countering illegal hate speech online**. 2016. Disponível em: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct. Acesso em: 21 jun. 2021.

12 BENTES, Ana. A gestão algorítmica da atenção: Enganchar, Conhecer e Persuadir In: Seminário Governança das Redes, 4., 2019, Belo Horizonte. **Políticas, Internet e Sociedade**. Belo Horizonte: IRIS. 984. p. 222-234. Disponível em: https://irisbh.com.br/publicacoes/politicas-internet-e-sociedade/. Acesso em: 13 maio 2021.

13 ZUBOFF, Shoshana. Big other: surveillance capitalism and the prospects of an information civilization. **Journal of information technology**, n. 30. Londres, 2015. Disponível em: https://cryptome.org/2015/07/big-other.pdf. Acesso em: 13 maio 2021

14 ANJOS, Lucas. **Testando os limites do capitalismo de vigilância**. 2019. Disponível em: https://irisbh.com.br/testando-os-limites-do-capitalismo-de-vigilancia/. Acesso em: 14 jun. 2021.

diversas anunciantes contra o escasso combate da plataforma ao discurso de ódio mobilizou a empresa a alterar abruptamente suas políticas de moderação[15].

- Opinião pública: ações coletivas propostas por pessoas usuárias ou grupos da sociedade civil já resultaram em alterações nas políticas de comunidade das plataformas. Este é o caso da modificação da política de moderação de fotos de amamentação após o protesto de mais de 80.000 mil mães quanto à restrição de informações relacionadas com o tema[16].

Os valores e interesses que motivam a moderação de conteúdo se revestem de formas concretas por meio da adoção de técnicas e práticas para intervenção no conteúdo gerado por terceiro. A subseção seguinte pretende aprofundar as principais formas de moderação de conteúdo utilizadas por intermediários.

1.2. Principais práticas e técnicas utilizadas para moderação de conteúdo

A motivação política para os esforços de moderação iniciou-se com o foco em diminuir a incidência de conteúdo danoso dentro das plataformas[17]. A veiculação desse tipo de conteúdo trazia prejuízos tanto de ordem humanitária, pois criava uma ambiente virtual hostil e permitia comunicações que violavam direitos humanos, quanto de ordem comercial, pois os patrocinadores das plataformas consideravam prejudicial associar-se a uma empresa que permitia tal forma de comunicação.

15 SANDOVAL, Pablo Ximénez. Cresce o boicote de grandes anunciantes contra o Facebook apesar da reação de Zuckerberg **El País**. 29 jun 2019. Disponível em: https://brasil.elpais.com/tecnologia/2020-06-29/cresce-o-boicote-de-grandes-anunciantes-contra-o-facebook-apesar-da-reacao-de-zuckerberg.html. Acesso em: 11 maio 2021.

16 SWENEY, Mark. Mums Furious as Facebook Removes Breastfeeding Photos. **The Guardian.** 30 Dez. 2008. Disponível em: https://www.theguardian.com/media/2008/dec/30/facebookbreastfeeding-ban [https://perma.cc/2Y2B-9HDA. Acesso em: 02 mar. 2021.

17 Um bom exemplo desses esforços políticos pode ser encontrado na Lei de Decência das Comunicações (*Communications Decency Act* - CDA) dos EUA: mais especificamente no item C de sua seção 230, que trata da chamada cláusula do "Bom Samaritano". Trata-se de um dispositivo que garante a invulnerabilidade de plataformas digitais perante conteúdo ilegal compartilhado por seus usuários, desde que essas plataformas promovam, de boa fé, a limitação ou restrição de conteúdo compartilhado em questão. O caso *Zeran vs. America Online (AOL)* é um dos mais paradigmáticos quanto a esse dispositivo legal, e reforçou a validade da seção 230 do CDA no Direito estadunidense. Para mais informações, confira: RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021. E também: KLONICK, Kate. The new governors: The people, rules, and processes governing online speech. **Harv. L. Rev.**, v. 131, p. 1598-1670, 2017. p. 1604-1609. Disponível em: <ttps://scholarship.law.stjohns.edu/cgi/viewcontent.cgi?article=1184&=&context=faculty_publications&=&sei-redir=1&referer=https%253A%252F%252Fscholar.google.com.br%252Fscholar%253Fhl%253Dpt-BR%2526as_sdt%253D0%25252C5%2526q%253Dkate%252Bklonick%2526btnG%253D#search=%22kate%20klonick%22. Acesso em: 10 fev. 2021.

Por essa razão, ao longo do tempo, governos, empresas e a sociedade civil têm se engajado para pressionar as plataformas a intervirem na redução de conteúdos danosos[18].

A amplitude do conceito de conteúdo danoso se dá não apenas porque a categoria abarca diversos tipos de conteúdo, mas também porque a própria ideia de "dano" é abrangente por si só. Para além da ideia de dano de um ponto de vista jurídico – de situação que enseja indenização perante um ordenamento –, o conceito de conteúdo danoso também abarca os conteúdos proibidos pelas próprias políticas internas das plataformas[19].

Um estudo[20] publicado pelo Berkman Klein Center for Internet & Society buscou analisar as práticas de moderação de conteúdo na versão em língua inglesa da plataforma Wikipédia. As modalidades de conteúdo utilizadas para compor o conceito de discurso danoso são oito, entre as quais pode-se observar grande intersecção: 1. assédio, bullying e *stalking*; 2. ameaças físicas e incitação de violência; 3. ataques baseados na identidade da vítima[21]; 4. *doxxing*[22]; 5. difamação; 6. desinformação; 7. vandalismo; e 8. discurso ilegal.

Entretanto, outras categorias de conteúdo também sofrem modulação em relação a seu alcance, visualização e circulação, como anúncios, conteúdo político e conteúdo protegido por propriedade intelectual. Portanto, desde já, destacamos que, ainda que a moderação de conteúdo danoso (ou potencialmente danoso) receba maior atenção midiática e regulatória, qualquer tipo de intervenção no conteúdo gerado por usuário deve ser realizada de forma transparente. Não obstante, em casos especiais, como assuntos relevantes para segurança nacional, prevenção a terrorismo, crimes internacionais, ou informaçõe sensíveis sobre o usuário, esse nível de transparência deve ser modulado para respeitar garantias e direitos. Devido à importância de se estender a atenção aos outros tipos de moderação de conteúdo, listamos a seguir uma gama de possibilidades de intervenção.

---

18 Como, por exemplo, a campanha "*Stop Hate for Profit*", à qual diversas marcas aderiram para demandar que plataformas exerçam maior controle sobre conteúdo odioso. Ver: **STOP Hate for Profit**. Disponível em: https://www.stophateforprofit.org/. Acesso em: 02 mar. 2021.

19 Nesse sentido, vale relembrar que as previsões legais sobre condutas ilícitas variam consideravelmente entre os diversos países abrangidos pelas plataformas digitais. Assim, é usual que as políticas internas dessas plataformas sejam criadas de forma a representar um denominador comum entre diversos regulamentos distintos, a fim de padronizar a experiência dos usuários e facilitar a administração do serviço em escala global.

20 CLARK, Justin et al. Content and Conduct: How English Wikipedia Moderates Harmful Speech. **The Social Science Research Network Electronic Paper Collection, Research Publication**, n. 2019-5, 2019. Disponível em: https://dash.harvard.edu/bitstream/handle/1/41872342/Content%20and%20ConductHow%20English%20Wikipedia%20Moderates%20Harmful%20Speech.pdf?sequence=1&isAllowed=y. Acesso em: 09 fev. 2021.

21 Tradução livre de "*Identity-based attacks*"

22 O termo *doxxing* diz respeito à divulgação de informações pessoais de alguém na internet, muitas vezes como forma de assédio, ameaça ou vingança. Ver: AVAST. **Doxxing: O que é e como evitá-lo**. Disponível em: https://www.avast.com/pt-br/c-what-is-doxxing. Acesso em: 17 jun. 2021.

a) Práticas de intervenção no conteúdo

No meio digital, algumas formas de moderação de conteúdo ganharam notoriedade devido à intensidade com a qual têm sido empregadas. Algumas dessas principais categorias são úteis para compreender as dimensões das intervenções sobre expressão online:

- Remoção[23]: consiste na eliminação completa de informação da plataforma, que passa a ser inacessível para todos os usuários. Recorrentemente, esse método de moderação é empregado com relação a conteúdo considerado impróprio ou ilícito pela plataforma, a fim de minimizar os danos decorrentes de sua veiculação ou posterior responsabilização de indenizar pessoas usuárias e terceiras lesadas.

- Indisponibilização: consiste em uma técnica por meio da qual um conteúdo deixa de ser acessível de forma limitada temporal ou geograficamente. Uma das situações em que se aplica esse método é quando o conteúdo em questão exige análise pormenorizada para ser veiculado amplamente – como por exemplo em casos de análise de alegadas infrações de propriedade intelectual ou de identificação automática de um potencial conteúdo danoso[24]. Outra possibilidade é quando existe alguma restrição legal àquele conteúdo em determinada jurisdição, o que leva esse conteúdo a ser inacessível em áreas específicas do globo – o chamado *geoblocking*[25].

- Restrição: assemelha-se às categorias apresentadas acima no sentido de limitar o alcance de uma informação veiculada através de uma plataforma digital. Pode ocorrer, por exemplo, por meio da ocultação do conteúdo com aviso sobre o teor e opção de vê-lo na forma completa ou não. Uma de suas formas é a restrição por idade (conteúdo impróprio para pessoas mais novas que determinada idade). Além disso, é comum encontrar conteúdo restrito em virtude de seu teor potencialmente traumático (violento ou perturbador, por exemplo).[26] Em

23 SRINIVASAN, Kumar Bhargav et al. Content removal as a moderation strategy: Compliance and other outcomes in the changemyview community. **Proceedings of the ACM on Human-Computer Interaction**, v. 3, n. CSCW, p. 1-21, 2019. p. 2-3. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3359265. Acesso em: 03 fev. 2021.

24 O YouTube, em sua página de ajuda quanto ao Content ID (ferramenta utilizada para detecção automática de supostas infrações a direitos autorais), apresenta o bloqueio do conteúdo como uma das opções para que o alegado detentor da propriedade intelectual em questão possa contestar um uso indevido de seu conteúdo. Para mais informações, ver: GOOGLE. **Como funciona o Content ID**. Disponível em: https://support.google.com/youtube/answer/2797370?hl=pt-BR#zippy=%2Cquais-opções-estão-dispon%C3%ADveis-aos-proprietários-de-direitos-autorais. Acesso em: 04 fev. 2021.

25 TRIMBLE, Marketa. Geoblocking, Technical Standards and the Law. In: LOBATO, Ramon; MEESE, James (org.) **Geoblocking and global video culture**. 2016. Disponível em: http://scholars.law.unlv.edu/facpub/947. Acesso em: 09 fev 2021.

26 As diretrizes de comunidade do YouTube, por exemplo, apresentam informações detalhadas sobre os referidos critérios de restrição de conteúdo por conteúdo impróprio para menores, violento, entre outras hipóteses. Para mais informações, confira: YOUTUBE. **Diretrizes da comunidade**. Disponível em: https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#community-guidelines.

plataformas que veiculam anúncios, pode ser impossibilitado à pessoa criadora auferir lucro com publicidade associada ao conteúdo ("desmonetização"[27]). Por isso, a medida representa um desincentivo à produção de determinados tipos de conteúdo.

- A sinalização: são avisos, anexados ao conteúdo em questão, que buscam alertar para publicações patrocinadas, de conteúdo eleitoral, ou ainda sobre a possibilidade de conter informações enganosas, imprecisas ou equivocadas. Esse método de moderação tem recebido grande atenção de redes sociais[28] em decorrência da crescente preocupação relativa ao compartilhamento de desinformação.

- Ranqueamento: uso de técnicas algorítmicas para determinar quais conteúdos serão priorizados em cada plataforma. Nesse sentido, há uma manipulação do que é considerado "relevante" ou não dentro de cada plataforma, dando maior visibilidade a informações melhor ranqueadas[29]. Embora se relacione à amplificação do alcance de determinado conteúdo, pode também servir para ofuscá-lo, limitando a possibilidade de destaque a alguns tipos de conteúdo, ou deixando em segundo plano informações que estejam desalinhadas com os parâmetros e interesses da plataforma[30].

Apesar de a maioria dos esforços relacionados à moderação de conteúdo em plataformas digitais referir-se a conteúdo danoso, é importante pontuar que todo tipo de conteúdo que circula nesses meios é objeto de moderação por parte das

Acesso em: 03 fev. 2021.

27 CAPLAN R.; GILLESPIE, T. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. **Social Media + Society**. Jun. 2020. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2056305120936636#articleCitationDownloadContainer. Acesso em: 3 fev. 2021.

28 A título de exemplo mais recente, pode-se citar também a sinalização de veracidade para alertar sobre publicações que alegavam fraudes no processo eleitoral presidencial dos Estados Unidos, em 2020. Também foram utilizados marcadores em postagens do Instagram sobre a pandemia de COVID-19, que apontam para links com informação confiável. Ver: G1. **Twitter e Facebook bloquearam páginas e postagens nesta campanha eleitoral nos EUA**; relembre casos. 15 out. 2020. Disponível em: https://g1.globo.com/mundo/eleicoes-nos-eua/2020/noticia/2020/10/15/twitter-e-facebook-bloquearam-paginas-e-postagens-nesta-campanha-eleitoral-nos-eua-relembre-casos.ghtml Acesso em: 18 jun. 2021. E ainda: INSTAGRAM. **Fornecer informação, segurança e apoio às pessoas no Instagram**. 24 mar. 2020. Disponível em: https://about.instagram.com/pt-br/blog/announcements/coronavirus-keeping-people-safe-informed-and-supported-on-instagram Acesso em: 18 jun. 2021

29 YALÇIN, Nursel; KÖSE, Utku. What is search engine optimization: SEO?. **Procedia-Social and Behavioral Sciences**, v. 9, p. 487-493, 2010. p. 488-489. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877042810022901>. Acesso em: 03 fev. 2021.

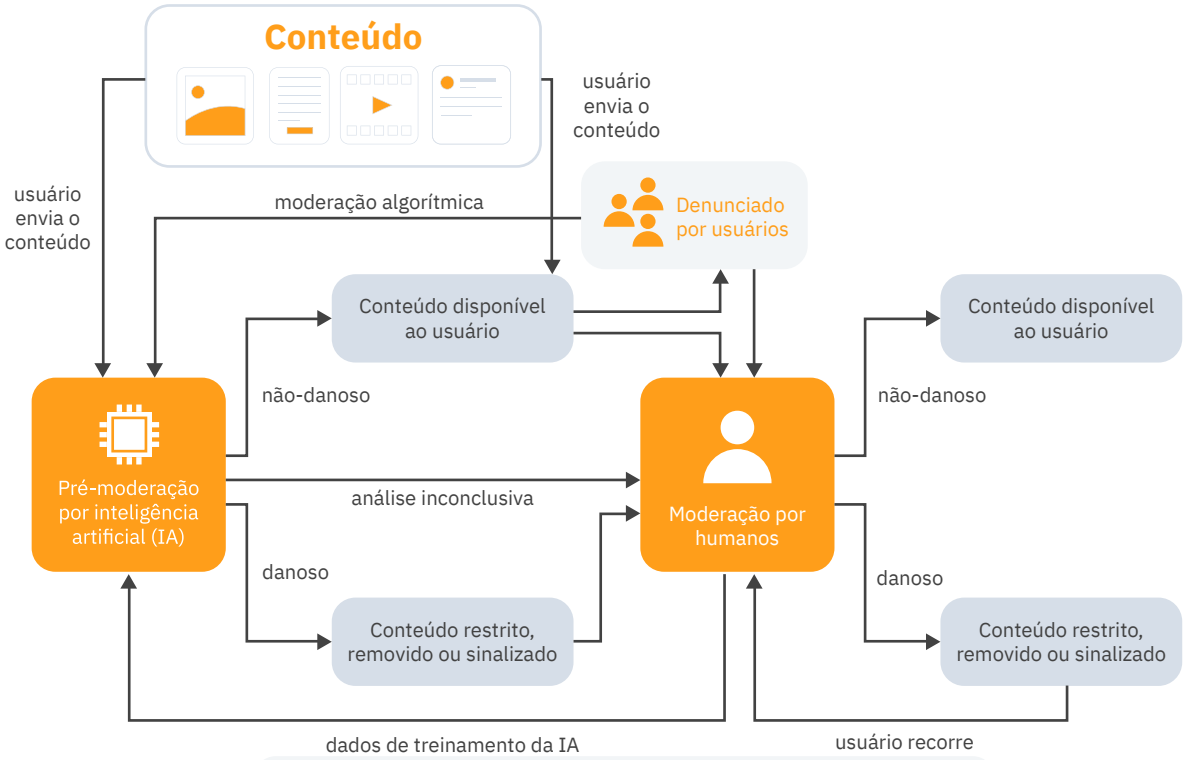
30 Essa prática é conhecida como *shadowban* ("banimento fantasma") quando os usuários não são informados que aquele tipo de conteúdo não será priorizado, porém, verifica-se que apesar de haver volume de postagens, determinados tópicos não são visibilizados em sugestões, páginas iniciais, buscas, etc. Ver mais em: MYERS WEST, Sarah. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. **New Media & Society**, v. 20, n. 11, p. 4366-4383, 2018. Ver também: INSTAGRAM. Central de Ajuda. Disponível em: https://help.instagram.com/613868662393739. Acesso em : 08 fev. 2021.

plataformas, inclusive conteúdo não danoso. Há diferentes maneiras de intervenção nos conteúdos em plataformas digitais, que também são englobadas no conceito de "moderação de conteúdo" apesar de serem aplicadas a informações plenamente legítimas, independentemente de causarem ou não dano. Alguns exemplos podem ser publicações de usuários em redes sociais, conteúdo jornalístico e notícias em geral, publicidade política, imagens, sites, entre outros.

Ainda, é possível que a moderação não seja exercida com relação a um conteúdo específico, mas sim a todo um perfil (ou usuário) que veicula esse conteúdo. É o caso, por exemplo, do banimento de contas em redes sociais, como o episódio em que a conta do Twitter do então presidente dos EUA, Donald Trump, foi suspensa permanentemente, em meio a uma série acusações de fraude na corrida eleitoral estadunidense de 2020[31].

Todo conteúdo, mesmo que legítimo, veiculado por meio de uma plataforma digital está sujeito a técnicas de moderação. Nesse sentido, vale ressaltar também que diferentes formas pelas quais essas decisões são tomadas, se por meio de análise humana ou automatizada, apontam para distintas questões a serem observadas na adequação ao direito de livre expressão.

b) Técnicas: Moderação algorítmica e humana



Adaptado do relatório feito pela Cambridge Consultants[32]

31 Para saber mais sobre a fundamentação da decisão de banimento da rede social,confira: TWITTER. Permanent suspension of @realDonaldTrump. **Twitter Blog**, 08 de janeiro de 2021. Disponível em: <https://blog.twitter.com/en_us/topics/company/2020/suspension>. Acesso em: 17 jun. 2021.

32 CAMBRIDGE CONSULTANTS. Use of AI in online content moderation. 2019. Disponível em: https://www.epra.org/news_items/ofcom-s-report-the-use-of-ai-in-content-moderation. Acesso em 13

A moderação de conteúdo pode ser exercida por diversas razões, com diversos objetivos e de variadas formas. Ainda assim, uma característica comum em meio a tantas variáveis é que a maior parte desse conteúdo moderado passa primeiro por uma análise automatizada feita com base em padrões algorítmicos e com uso de inteligência artificial[33].

É bastante frequente a prática de dois tipos de análise prévia: o ranqueamento e a filtragem. A primeira forma é aquela na qual o algoritmo decide quais conteúdos serão entregues a cada usuário, por meio de análise automatizada de suas preferências e outros dados de uso. Algoritmos e inteligência artificial também são utilizados para o crescimento comercial das plataformas, para a manutenção do engajamento da pessoa na plataforma, para direcionamento de conteúdo, para disseminação de determinada ideia, entre outras possibilidades.

A outra forma de moderação automatizada consiste na inspeção de conteúdo publicado, em busca de padrões que indiquem que ele pode ser indesejado naquela plataforma e sua circulação representa um risco[34]. A partir dela, o algoritmo pode aplicar ou apontar para algum tipo de medida interventiva – desvalorização no ranqueamento, *shadowban*[35], remoção, suspensão, a sinalização, por exemplo.

Devido aos diversos fatores que envolvem a tomada de decisão na moderação, a análise de conteúdo por humanos é essencial e faz parte do processo de moderação, especialmente em zonas cinzentas para os algoritmos. Entretanto, considerando que o trabalho de moderação de conteúdo muitas vezes é analisar publicações traumatizantes – por serem violentas ou asquerosas, entre outras – o prejuízo dessa atividade à saúde mental dos profissionais é pungente, como exposto por pesquisadores[36], pelas próprias pessoas que trabalham com isso[37] e reconhecido,

maio 2021

33 CAMBRIDGE CONSULTANTS. **Use of AI in online content moderation**. 2019. Disponível em: https://www.epra.org/news_items/ofcom-s-report-the-use-of-ai-in-content-moderation. Acesso em 13 maio 2021.

34 Esse tipo de moderação se aplica a padrões visuais ou escritos que lidem com temas sensíveis ou ilícitos, como, por exemplo, violações a propriedade intelectual, crimes violentos, pornografia infantil, e, em alguns casos, nudez, discurso de ódio, substâncias ilícitas, etc.

35 Técnica por meio da qual a plataforma reduz o alcance de certos conteúdos sem que os usuários que o publicam tenham ciência da redução. Ver: RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021. p. 73

36 STEIGER, Miriah et al. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support . **CHI '21:** Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, maio 2021. Disponível em: https://crowd.cs.vt.edu/wp-content/uploads/2021/02/CHI21_final__The_Psychological_Well_Being_of_Content_Moderators-2.pdf . Acesso em 14 jun. 2021.

37 WAKEFIELD, Jane. 'Após trabalhar como moderadora na web, parei de apertar a mão das pessoas – fiquei com nojo da humanidade'. **BBC News**. 15 out. 2018. Disponível em: https://www.bbc.com/portuguese/geral-45861208. Acesso em 14 jun. 2021.

em alguma medida, pelas plataformas[38].

A pesquisadora Sarah Roberts relata que pessoas que trabalham na área frequentemente são contratadas como autônomas, em tempo parcial, sem o status das empregadas regulares. Ainda, é comum que essa mão-de-obra seja terceirizada para empresas em países onde há menor custo na contratação, flexibilização de direitos trabalhista e precarização da atividade profissional[39]. Embora a moderação de conteúdo por humanos seja vista como um fator necessário e recomendado em documentos de referência como os Princípios de Santa Clara[40], sobretudo em procedimentos específicos, como o de revisão de decisões, ela enfrenta limitações e críticas. Tanto moderação algorítmica quanto humana são necessárias ao funcionamento de plataformas de conteúdo gerado por usuário. Por isso, existem questões regulatórias e tensões quanto aos limites da atuação de ambos.

Esses são alguns exemplos de práticas adotadas para se realizar a moderação de conteúdo em plataformas digitais e os dilemas nelas envolvidos. Pode-se perceber que se trata de uma disciplina que está submetida a diversas variáveis, relacionadas tanto à natureza das informações veiculadas quanto ao contexto regulatório aplicável às plataformas nos âmbitos local e internacional.

1.3. A transparência na moderação de conteúdo

Esta seção pretende explorar como a transparência e a moderação de conteúdo se relacionam, além de compreender os pontos de análise a serem considerados nos esforços de transparência. Isso será feito tanto para explicitar quais os conceitos utilizados neste trabalho quanto para destacar que a transparência não é uma solução total ou única para as questões levantadas.

1.3.1. O que é transparência?

Transparência refere-se a qualquer prática que torna informações internas a uma organização acessíveis e visíveis a atores fora do âmbito de produção da informação, ou seja, refere-se à abertura na transmissão de informações[41]. O grau de abertura e

38 ESTADÃO. Facebook pagará US$ 52 mi para funcionários com problemas na saúde mental. **Exame.** 14 maio 2020. Disponível em: https://exame.com/negocios/facebook-pagara-us-52-mi-para-funcionarios-com-problemas-na-saude-mental/. Acesso em 14 jun. 2021.

39 ROBERTS, Sarah T. Digital detritus:'Error'and the logic of opacity in social media content moderation. **First Monday.** 2018.

40 EFF et al. **Santa Clara Principles on transparency and accountability in content moderation**. Disponível em: https://santaclaraprinciples.org/. Acesso em: 12 fev. 2021.

41 FINEL, Bernard I.; LORD, Kristin M. 1999. The Surprising Logic of Transparency. **International Studies Quarterly** 43, no. 2, p. 316. Disponível em: https://transparencylab.org/Documentation/Additional%20resources/Literature/Finel%20et%20al_The%20Surprising%20Logic%20of%20

o contexto das informações que são visibilizadas são variáveis relacionadas ao nível de transparência exercido.

Transparência comumente é uma característica relacionada à administração pública ou ao governo. De fato, no início dos debates teóricos sobre o assunto, marcado por Jeremy Bentham no século XIX[42], os apelos para a publicação de contas públicas e encargos centralizavam grande parte dos debates sobre transparência. Entretanto, com os novos arranjos sociais, o conceito de transparência agora alcança outras formulações, sendo menos vinculado à figura do Estado e mais contrastado com medidas geralmente opacas. Nessas situações é difícil descobrir quem toma as decisões, quais são elas, quem ganha e quem perde com elas[43].

A disponibilização de informações em direção à transparência é almejada por potencialmente promover um melhor desempenho em organizações e indivíduos, seja pelo aumento da racionalização da sociedade, seja pelo desestímulo a comportamentos considerados negativos, como a corrupção. A transparência, nesse sentido, é muitas vezes vista como um elemento necessário para manutenção da confiança entre uma entidade e o público geral. É possível observar a atribuição de maior importância aos instrumentos de transparência conforme aumenta o abismo hierárquico-informacional entre as partes comunicadora e receptora da informação. Assim, para diminuir as assimetrias, divulga-se a informação e fomenta-se a confiabilidade da relação.

1.3.2. Relação com o direito à liberdade de expressão

A relação entre transparência e moderação de conteúdo é interligada pela conexão de ambas com o direito à liberdade de expressão. O argumento central é que a moderação de conteúdo interfere nas manifestações individuais dos cidadãos – e altera holisticamente as próprias condições sociais do discurso[44]. Segundo essa lógica, devido à centralidade do discurso na concretização da democracia, qualquer forma de interferência, manipulação ou controle deve ser acompanhada de perto. Logo, a transparência é uma prática aliada à promoção da democracia.

Transparency%20_1999.pdf. Acesso em 06 mar. 2021.

42 BIRCHALL, Clare. Introduction to 'Secrecy and Transparency' The Politics of Opacity and Openness. **Theory, Culture & Society**, v. 28, n. 7-8, p. 7-25, 2011. Disponível em: https://journals.sagepub.com/doi/10.1177/0263276411427744. Acesso em 14 jun. 2021.

43 BLACK, Julia. Transparent policy measures. **Oxford Dictionary of Economics**, Oxford University Press, Oxford, 1997. Disponível em: https://www.oxfordreference.com/view/10.1093/acref/9780199237043.001.0001/acref-9780199237043-e-3191. Acesso em 13 maio 2021.

44 BALKIN, Jack M., Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society. **New York University Law Review**, Vol. 79, No. 1, 2004, Yale Law School, Public Law Working Paper No. 63, Disponível em: https://ssrn.com/abstract=470842. Acesso em 11 maio 2021.

A primeira premissa – de que a moderação de conteúdo interfere na liberdade de expressão do usuário – é evidenciada por Gillespie, no sentido que as plataformas buscam direcionar o usuário para a experiência "certa" de navegação online.

> As plataformas usam essas três alavancas juntas para, ativa e dinamicamente, ajustar a participação inesperada dos usuários, produzir o feed "certo" para cada usuário, as trocas sociais "certas", o tipo "certo" de comunidade. ("Certo" aqui pode significar ético, legal e saudável; mas também significa qualquer coisa que promova o engajamento, aumente a receita de anúncios e facilite a coleta de dados)[45].

Apesar de não resultar necessariamente em censura, a moderação de conteúdo deve ser realizada sob a ótica da transparência e devido processo, pois é uma interferência na liberdade de expressão. Ela é feita, como já mencionado, com base nos interesses comerciais, políticos e econômicos da plataforma. A moderação não é feita pelas plataformas de forma neutra, tampouco considerando o interesse do usuário como principal objetivo.

O impacto da moderação de conteúdo na liberdade de expressão é relevante a ponto de as principais empresas que realizam moderação serem chamadas de "governadoras do discurso online"[46], "guardiões da esfera pública"[47] e "administradoras da cultura pública"[48]. As plataformas integram o cenário globalizado[49], uma vez que são, via de regra, corporações com usuários e negócios internacionais. Os limites que impõem à circulação ou à publicação de informação por seus usuários têm reflexos tanto sobre o acesso à informação quanto sobre a liberdade de expressão.

Alguns exemplos ilustram os riscos envolvidos em reforçar ou excluir ideias pela moderação de conteúdo. Recentemente, em 2020, vieram a público, por meio de uma reportagem[50], regras internas de remoção de conteúdo da plataforma TikTok,

45 GILLESPIE, Tarleton. **Custodians of the Internet**: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press, 2018. p. 217, p. 18

46 KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. **Harvard Law Review**, v. 131. n. 6. p. 1598-1670, abril de 2018. Disponível em: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Acesso em: 17 jun. 2021. Acesso em 18 jun. 2021. p. 1625-1630

47 BALKIN, Jack M. Free speech is a triangle. **Columbia Law Review**, v. 118, p. 2011, 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3186205 Acesso em: 18 jun. 2021.

48 GILLESPIE, Tarleton. Platforms are not intermediaries. **Georgetown Law Technology Review**, v. 2, n. 2, p. 198-216, 2018. Disponível em: https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Gilespie-pp-198-216.pdf Acesso em: 18 jun. 2021.

49 MICHOUD, Adeline. Corporate Accountability in International Law: A'State Shaming'Exercise. **SOAS Law Journal**, v. 7, 2020. Disponível em: https://8c607a4d-3252-415f-9696-9203fde0b41e.filesusr.com/ugd/79f2dc_ce5b501dbccc49f98cc6b3ae35d59830.pdf Acesso em: 18 jun. 2021. p. 190.

50 BIDDLE, Sam; RIBEIRO, Paulo Victor; DIAS, Tatiana. Censura invisível. Exclusivo: TikTok

consideradas ofensivas aos direitos humanos. Entre elas, eram proibidas diversas formas de expressão por pessoas ou em ambientes que não pertencessem a classes sociais privilegiadas ou idealizadas como padrão de beleza. Havia, além disso, uma instrução nos documentos para a exclusão de publicações sobre a independência do Tibet, o que representa uma barreira à livre informação e debate sobre assuntos políticos. Há também o algoritmo de corte de imagem da plataforma Twitter, denunciado pela comunidade de usuários por priorizar pessoas brancas sobre negras, mesmo quando elas não seriam o destaque[51]. Também foi amplamente divulgado o caso da modelo Nyome Nicholas-Williams, mulher negra que teve suas fotos excluídas pelo Instagram, cuja política considerava como violação a exibição parcial de seios de mulheres que não fossem consideradas magras[52]. A repercussão de todos esses eventos levou à adoção de medidas e ajustes nas práticas de moderação de conteúdo[53]. Isso demonstra a relevância de permanente discussão e reavaliação das medidas de moderação de conteúdo.

Plataformas são vistas como agentes relevantes e discutem-se formas de aliar essa atuação ao cenário internacional de direitos humanos. A maior influência sobre moderação de conteúdo atual, a seção 230 do *Communications Decency Act* (CDA), é objeto de questionamentos e propostas de revisão. O CDA é a lei estadunidense que estipula o modelo de autorregulação de intermediários sobre conteúdo na internet, com uma dupla proteção contra responsabilização, da seguinte forma: as plataformas podem moderar conteúdo de seus usuários conforme considerarem compatível com suas políticas, sem serem responsabilizadas pela manutenção ou retirada do conteúdo. Isso significa que não há o dever das intermediárias de indenizar a pessoa ofendida por conteúdo danoso mantido nem por remoção indevida de conteúdo legítimo. Essa legislação faz parte do contexto de incentivo tecnológico e de desenvolvimento das plataformas online. Busca, ainda, resguardar a liberdade de expressão e evitar a imposição de censura colateral[54]. Isso ocorre quando a remoção

censurou rostos feios e favelas para atrair novos usuários e censurou posts políticos. **The Intercept Brasil**. 16 mar. 2020. Disponível em: https://theintercept.com/2020/03/16/tiktok-censurou-rostos-feios-e-favelas-para-atrair-novos-usuarios/ Acesso em: 12 fev. 2021.

51 CARVALHO, Caio. Após acusações de racismo, Twitter promete atualizar algoritmo de corte de imagens. **Gizmodo Brasil**. 2 out. 2020. Disponível em: https://gizmodo.uol.com.br/twitter-atualizar-algoritmo-corte-imagens-racismo/. Acesso em: 12 fev. 2021.

52 TATSCH, Constança. Gordofobia: Instagram muda regras sobre exibição de seios após protesto. **Extra**. 26 out. 2020. Disponível em: https://extra.globo.com/noticias/gordofobia-instagram-muda-regras-sobre-exibicao-de-seios-apos-protesto-24713044.html. Acesso em: 11 maio 2021.

53 Por exemplo, o Twitter constatou que seu algoritmo discriminava mulheres e pessoas negras e, por isso, deixou de usar o mecanismo de corte de imagens. LISBOA, Alveni. Twitter constata que seu algoritmo de corte de fotos é mesmo racista. **Canaltech**, 20 maio 2021. Disponível em: https://canaltech.com.br/apps/twitter-constata-que-seu-algoritmo-de-corte-de-fotos-e-mesmo-racista-185388/. Acesso em: 09 jun 2021.

54 KLONICK, Kate. The new governors: The people, rules, and processes governing online speech. **Harvard Law Review**, v. 131, 2017. Disponível em: https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/. Acesso em 18 jun. 2021. p. 1598.

de conteúdo pela plataforma torna-se padrão, por receio de uma condenação judicial, inclusive em casos nos quais o conteúdo seria legítimo.

Um fator para o protagonismo do CDA é que as principais plataformas e empresas que oferecem serviços na internet – como Google, Amazon, Facebook, Apple, Microsoft (que formam o acrônimo popularizado como GAFAM) – são sediadas nos Estados Unidos da América[55]. Uma consequência disso é que as equipes que redigiram as políticas contam com liderança por advogados especializados na Primeira Emenda à Constituição daquele país, que alça a liberdade de expressão ao patamar de prioridade[56].

A censura colateral, entretanto, não é o único risco a ser evitado no cenário da moderação de conteúdo. A ideia de uma regulação material, isto é, que imponha diretrizes sobre o conteúdo a ser moderado diretamente, também levanta preocupações, tanto no setor privado quanto na sociedade civil[57]. Isso abriria margem para regras demasiadamente restritivas, que interferem no acesso à informação de todos os usuários. Ainda, a regulação material poderia contribuir com um processo de fragmentação da rede[58], com cada local operando conteúdos diversos a depender da jurisdição, o que reduz o potencial comunicativo global da internet.

Em episódio recente, essa discussão veio à tona na mídia[59] devido à suspensão da conta do Twitter do então presidente dos Estados Unidos, Donald Trump. A plataforma, após já ter suspendido a conta do ex-presidente temporariamente, decidiu pela suspensão permanente em razão de suas postagens de apoio à invasão violenta do Capitólio em protesto ao resultado das eleições de 2020, em que Joe

55 HELDT, Amélie Pia. Reading between the lines and the numbers: an analysis of the first NetzDG reports. **Internet Policy Review**, v. 8, n. 2, 2019. Disponível em: https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports. Acesso em: 18 jun. 2021. p. 340

56 KLONICK, Kate. op. cit., 2017.

57 FAGAN, Frank. Systemic social media regulation. **Duke Law & Technology Review**, v. 16, p. 393, 2017. Disponível em: https://scholarship.law.duke.edu/dltr/vol16/iss1/14/ Acesso em: 18 jun. 2021.

58 "O processo de fragmentação da internet por mecanismos técnicos e jurídicos tem sido chamado de 'balcanização' da internet. O termo faz referência à fragmentação política dos Estados do Sul da Europa, em razões de diferenças étnicas, religiosas e culturais, após o fim do domínio estrangeiro sobre a região. O fenômeno se caracteriza quando programas governamentais de censura, interesses comerciais, preocupações com cibersegurança e outras mudanças dinâmicas no ecossistema da internet acabam por retalhar a rede global em diversas versões regionais. Essa retaliação ameaça a comunicação universal, a inovação e a prosperidade econômica trazida pela internet como inicialmente estruturada." Citação de ANTUNES, Laila Damascena; ROSA, Matheus; VILELA, Pedro. **Jurisdição e internet:** Estudo sobre mecanismos de bloqueio e fragmentação da rede. Instituto de Referência em Internet e Sociedade: Belo Horizonte, 2017. Disponível em: http://bit.ly/2PjUqYT. Acesso em: 14 jun. 2021.

59 GOLDMAN, Alex; VOGT, P.J. **Reply All**. Account Suspended. Disponível em: https://gimletmedia.com/shows/reply-all/94h3el6/171-account-suspended Acesso em: 03 fev. 2020.

Biden foi escolhido como novo presidente[60]. O argumento da empresa foi o risco das manifestações de Trump incitarem ainda mais violência, prática proibida nas políticas de comportamento do Twitter[61]. Por ser uma ação de interferência na liberdade de expressão de um líder político notório, a repercussão da medida foi global e múltiplas perspectivas foram apresentadas[62] [63] e ilustram a complexidade da questão.

A transparência desempenha papel relevante por permitir discutir concretamente os reflexos da moderação de conteúdo sobre a livre expressão, sejam eles diretos ou colaterais.

1.3.3. Importância da transparência algorítmica

Os interesses das plataformas manifestam-se de diferentes formas no processo de moderação de conteúdo. Por exemplo, é possível que as políticas de comunidade[64] – que em resumo são as regras elaboradas pela plataforma para boa convivência das pessoas na rede – sejam construídas para estimular a publicação de determinados conteúdos em detrimentos de outros. Além disso, na própria prática de moderação de conteúdo são observados interesses de diferentes segmentos determinando o que permanece online e o que é limitado. O enviesamento algorítmico é um exemplo concreto dessa forma de interferência velada.

A definição mais comum de algoritmos entende-os como modelos matemáticos construídos por programadoras e programadores a partir de uma sequência de ações que tem por consequência uma ação esperada[65]. Aqui, o primeiro questionamento

60 NEWTON, Casey. **The Platformer**. Twitter Dumps Trump. Disponível em: https://www.platformer.news/p/-twitter-dumps-trump Acesso em: 03 fev. 2020.

61 TWITTER. **Central de ajuda**. Disponível em: https://help.twitter.com/pt/rules-and-policies/twitter-rules Acesso em: 11 maio 2021.

62 O próprio presidente do Twitter afirmou que a suspensão " estabelece um precedente que acredito ser perigoso: o poder que um indivíduo ou empresa tem sobre uma parte da conversa pública global" Ver: CLAYTON, James. Por que fundador do Twitter diz que banimento de Trump foi 'correto, mas perigoso'. **BBC**, 15 jan. 2021. Disponível em: https://www.bbc.com/portuguese/internacional-55674897 Acesso em: 11 maio 2021.

63 Além disso, o debate midiático foi fomentado tanto por críticos à medida da rede social, como o historiador Diego Lagedo que denominou a suspensão de "caça às bruxas modernas", quanto por apoiadores que como o advogado Thiago Lopes que afirmou que "O banimento de Trump do Twitter, por outro lado, não é censura, é reposicionamento corporativo." Ver: LAGEDO, Diego. 'Censura' a Donald Trump atenta contra a liberdade de expressão. **Blog de Jamildo**. 9 jan. 2021. Disponível em: https://blogs.ne10.uol.com.br/jamildo/2021/01/09/censura-a-donald-trump-atenta-contra-a-liberdade-de-expressao-por-diego-lagedo/ Acesso em: 11 maio 2021.

64 RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021

65 PAES, Natiele. Algorítimos: o que são e como fazem parte do cotidiano? **Politize**. 7 maio 2020. Disponível em: https://www.politize.com.br/algoritmos-o-que-sao/. Acesso em 11 maio 2021.

aparece: o enviesamento dos resultados devido à parcialidade dos programadores e até mesmo das bases de dados utilizadas para treinar os algoritmos.

Diversos casos de enviesamento ilustram a dimensão do problema. Em pesquisa realizada pelo Notícia Preta em 2019, uma busca no Google por "homens bonitos" resultou apenas cinco negros entre os cem primeiros resultados, enquanto que "mulheres bonitas" resultou apenas seis negras. Entretanto, esses números aumentaram para doze e treze, respectivamente, ao mudar os termos para "homens feios" e "mulheres feias". Outra pesquisa, realizada pelo InternetLab, identificou que o discurso de drag queens no Twitter é frequentemente associado a conteúdo danoso por ser descontextualizado da linguagem da comunidade LGBTQIA+ e ser considerado tóxico[66].

É preciso ter em conta que os preconceitos reforçados por algoritmos são frequentemente uma reprodução de preconceitos enraizados na sociedade. Assim, a discriminação manifestada por algoritmos não é um problema estrito do uso de tecnologias. Entretanto, existem especificidades, como, por exemplo, a escala com que ela se aplica e também a dificuldade em identificar qual o problema para corrigir o viés. Nesse sentido, é importante garantir a diversidade na equipe de programadores para que a automatização da moderação reconheça seus vieses e seja mais plural e contextualizada. Além disso, é necessário adotar medidas de transparência sobre as inferências e ações tomadas pelas máquinas de moderação a fim de concretizar os objetivos principais desta ação, como já mencionado: aumento da racionalização, melhoria na tomada de decisões e acompanhamento para limitação de abusos e condutas indesejadas.

Um dos caminhos para aumentar a transparência na moderação automatizada é a viabilização de auditoria dos sistemas e métodos utilizados pelas plataformas, como recomendado pelo AI Now Institute[67]. A auditoria de sistemas deve possibilitar que especialistas acessem sistemas de acordo com o nível e tipos de riscos envolvidos nas decisões automatizadas. **Além disso, a participação dos grupos afetados pelos potenciais riscos, em conjunto com especialistas e cientistas, é essencial para assegurar que tais comunidades tenham voz na tomada de subsídio e na formulação de questões.** Apesar das preocupações com a realização de auditorias por profissionais externos em razão de um potencial interesse econômico na questão,

66 Conforme o estudo, "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". GOMES, Alessandra; ANTONIALLI, Dennys; OLIVA, Thiago Dias. Drag queens e Inteligência Artificial: computadores devem decidir o que é 'tóxico' na internet? **InternetLab**. 28 jun. 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 11 maio 2021.

67 REISMAN, Dillon et al. Algorithmic impact assessments: A practical framework for public agency accountability. **AI Now Institute**, p. 1-22, 2018. Disponível em: https://ainowinstitute.org/aiareport2018.pdf. Acesso em 13 maio 2021.

conforme recomendação do AI Now Institute, uma entidade independente pode ser criada para supervisão desta auditoria nos limites da eticidade e legalidade.

Considerando os avanços historicamente conquistados para o reconhecimento e efetivação da liberdade de expressão como direito fundamental, a atenção a potenciais violações deste direito por meio da moderação de conteúdo – seja em razão de aplicação de regras arbitrárias, parcialidade na execução de programas ou descontextualização da tomada de decisões – reflete preocupações para que o uso das tecnologias da informação e comunicação esteja em consonância com modelos centrados no pleno desenvolvimento e respeito dos cidadãos e da coletividade.

1.3.4. Direito à privacidade e proteção de dados

Além da relação direta com o direito à liberdade de expressão, moderação de conteúdo e transparência estão intimamente relacionadas com o tratamento de dados pessoais. Especialmente em um contexto de moderação automatizada, a alta capacidade tecnológica dos softwares e hardwares envolvidos na moderação possibilita a extração e análise de detalhes profundos sobre a vida de um indivíduo[68]. Essa alta capacidade de tratamento de dados possibilita uma maior sincronia entre o que o usuário tem interesse em encontrar na plataforma e o que a plataforma lhe disponibiliza. Entretanto, a ausência de informação sobre quais são os dados coletados e analisados para a moderação de conteúdo impõe riscos à autodeterminação informativa do usuário. Os novos marcos normativos que trazem um arcabouço mais robusto para a proteção da privacidade do usuário, como a Lei Geral de Proteção de Dados[69] e o Regulamento Geral sobre a Proteção de Dados[70], buscam sanar esta lacuna.

Em nota técnica pública, o Laboratório de Políticas Públicas e Internet[71] destacou as principais inovações relacionadas à transparência trazidas pelas supracitadas regulamentações:

68 SILVA, Felipe Rocha et al. **NOTA TÉCNICA PL 2.630/2020** sobre a inclusão de mecanismos de transparência algorítmica no projeto de lei. LAPIN, 2020. Disponível em: https://lapin.org.br/wp-content/uploads/2020/08/LAPIN-NT-PL-2630-Transparencia-Algoritmica.pdf Acesso em 11 maio 2021.

69 BRASIL. Lei nº 13709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União.** Brasília, DF, 14 ago. 2018. Dísponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm Acesso em 11 maio 2021

70 UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). **Jornal Oficial da União Europeia.** 4 maio 2016. Disponível em: https://gdpr-info.eu/ Acesso em: 11 maio 2021.

71 SILVA, Felipe Rocha et al. **NOTA TÉCNICA PL 2.630/2020** sobre a inclusão de mecanismos de transparência algorítmica no projeto de lei. LAPIN, 2020. Disponível em: https://lapin.org.br/wp-content/uploads/2020/08/LAPIN-NT-PL-2630-Transparencia-Algoritmica.pdf. Acesso em 11 maio 2021.

> A LGPD já pressupõe mecanismos de transparência ativa, pela qual o titular de dados tem o direito de requerer "informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada" (art. 20, §1º). No entanto, falta a previsão de uma transparência passiva, pela qual o controlador de dados (como é o caso de uma plataforma) informa o usuário sobre a existência de decisões automatizadas e como elas ocorrem. Afinal, se a pessoa nunca é informada passivamente que decisões automatizadas estão ocorrendo, como ela vai tomar a iniciativa de solicitar alguma revisão delas?
>
> O Regulamento Geral de Proteção de Dados europeu - RGPD traz regras nesse sentido. Em seus artigos 13(2)(f) e 14(2)(g), a norma determina que as seguintes informações devem ser prestadas de forma espontânea pelo controlador de dados, independentemente de requisição: 1. A existência de decisões automatizadas; 2. Informações relevantes sobre a lógica das tomadas de decisão automatizadas; 3. Impacto e possíveis consequências dessas tomadas de decisão.

É essencial que os mecanismos e técnicas de transparência empregados por plataformas digitais em suas atividades de moderação de conteúdo sejam realizados em observância aos direitos dos usuários à privacidade e à proteção de dados pessoais[72]. Há considerações importantes que as plataformas devem fazer com relação a essa matéria. Não raramente, afinal, a transparência é atrelada – ao menos em certa medida – à sistematização de dados sobre os autores de conteúdos submetidos a moderação.

É o que se pode observar, em especial, em relatórios de transparência e em informações sobre a autoria de denúncias sobre a ilegalidade de um conteúdo, repassadas para a pessoa que publicou o material alegadamente impróprio em um primeiro momento. Garantir elementos de transparência mais efetiva em atividades dessa natureza pode acarretar o compartilhamento de dados pessoais de usuários com terceiros – o que, por sua vez, gera preocupações relevantes. Essa questão pode ser exemplificada pelo caso emblemático envolvendo a Cloudflare, em 2017[73]. A plataforma, que presta serviços de hospedagem de sites, compartilhou com seus clientes os dados – nomes e endereços de e-mail – de usuários que denunciaram

72 SUZOR, Nicolas P. et al. What do we mean when we talk about transparency? Toward meaningful transparency in commercial content moderation. **International Journal of Communication**, v. 13, p. 1526-1543, 2019. Disponível em: https://ijoc.org/index.php/ijoc/article/view/9736. Acesso em: 18 jun. 2021. p. 1537.

73 SCHWENKE, Ken. How One Major Internet Company Helps Serve Up Hate on the Web. **Propublica**. 4 maio 2017. Disponível em: https://www.propublica.org/article/how-cloudflare-helps-serve-up-hate-on-the-web. Acesso em: 18 jun. 2021.

postagens realizadas em seus respectivos sites. Isso significou, contudo, que administradores de sites ligados à propagação de conteúdo neonazista fossem informados em detalhe sobre esses dados, vinculados a pessoas que inequivocamente haviam manifestado resistência a conteúdo dessa natureza. Não tardou para que esses administradores compartilhassem publicamente os dados pessoais em sua posse, o que, por sua vez, resultou em ameaças aos autores das denúncias pelos frequentadores e apoiadores desses sites. Em resposta a esse evento, a empresa alterou sua política de denúncia de conteúdo considerado impróprio, e permitiu que em alguns casos essas denúncias sejam realizadas de forma anônima[74].

A situação supramencionada exemplifica a sensibilidade do processo de ampliação das técnicas de transparência empregadas por plataformas. Trata-se de uma questão na qual é imperativo haver um equilíbrio entre o objetivo propriamente dito – uma transparência efetiva – e a garantia da privacidade dos atores envolvidos, sob o risco de que os dados empregados no processo sejam utilizados por atores mal intencionados.

Em consonância com as razões apontadas anteriormente que tornam a transparência na moderação de conteúdo uma prática a ser almejada pela sociedade em geral, e praticada especialmente pelas plataformas, existem diversos elementos que caracterizam a transparência como significativa e efetiva. Percebe-se que certos tipos de transparência revelam informações, mas nem toda informação promove confiança. Para além do combate à ausência de informação ou da manutenção do segredo, é preciso superar ilusões ou enganos criados pelo excesso de informação ou por uma fachada de informação[75].

1.4. Uma perspectiva crítica sobre a transparência

Conforme teorizado por Flyverbom[76], dependendo da prática aplicada, os resultados dos esforços de transparência podem significar mais "gestão de visibilidade" do que discernimento e clareza. Ou seja, ao criar visibilidade e possibilidades particulares de observação, algumas partes da informação tornam-se passíveis de conhecimento e governabilidade, enquanto outras partes são ocultadas.

74 SCHWENKE, Ken. Internet Company That Does Business With Hate Sites Alters Complaint Policies. **Propublica**. 10 maio 2017. Disponível em: https://www.propublica.org/article/internet-company-does-business-with-hate-sites-alters-complaint-policies. Acesso em: 18 jun. 2021.

75 O'NEILL, Onora. **A question of trust:** The BBC Reith Lectures 2002. Cambridge University Press, 2002. p. 72

76 FLYVERBOM, Mikkel. Digital age| transparency: Mediation and the management of visibilities. **International Journal of Communication**, v. 10, p. 13, 2016. Disponível em: https://ijoc.org/index.php/ijoc/article/view/4490. Acesso em: 18 jun. 2021.

A divulgação de relatórios de transparência pelas plataformas pode se enquadrar nessa situação. Ainda que essenciais como ferramenta de maximização da disponibilização de informação acerca da interferência das empresas sobre conteúdos, tais relatórios são apresentados de forma unilateral pelas plataformas. Por esta razão, a interpretação dos relatórios é feita por pessoas e entidades sem forma alguma de diálogo institucional com a plataforma. Muitas vezes, tais documentos apenas apresentam as estatísticas relativas à moderação sem oferecer as bases necessárias para que o leitor compreenda o que elas significam. Dessa forma, um esforço para aumentar a transparência sobre a moderação de conteúdo pode resultar na disseminação de desinformação devido à publicação de dados isolados e descontextualizados[77] [78].

A concepção de transparência está intimamente ligada ao poder, pois um grupo ou indivíduo detém autoridade para decidir quem pode observar quem, quais são as atividades abertas e fechadas e quais são os objetos submetidos a esforços de transparência. Assim, críticas à possível descontextualização, seletividade de informações e as formas unidirecionais de transparência subsidiam correntes teóricas que questionam a concepção messiânica da transparência. A enxurrada de informação não classificada e o "massageamento da verdade"[79] que resulta em desinformação pode fazer com que a transparência corroa a confiança, conforme teorizado por O'Neill[80].

Para que os esforços de transparência não tenham efeitos indesejáveis ou promovam a ampliação de assimetrias de poder, é importante ressaltar os elementos que constroem a ponte entre "ver" e "mudar", conforme considerado por Annany e

77 Por exemplo, no relatório de transparência de 2020 do Facebook, o México aparece como o país com mais requisições para retirada de conteúdo por violar a lei local, tendo 6.335 requisições. Considerando que cerca de 200 países não tiveram nenhuma requisição para restrição de conteúdo com este motivo, de acordo com o relatório, leituras que consideram apenas o relatório de transparência podem inferir que os usuários do Facebook mexicanos cometem mais ilegalidades na plataforma. Entretanto, o relatório não informa as leis que supostamente foram violadas, o autor da requisição ou qualquer outra variável que possibilite a confirmação dessa hipótese, caracterizando-se como uma informação isolada que não resulta em transparência significativa. Ver: FACEBOOK. **Content Restrictions Based on Local Law.** Disponível em: https://transparency.facebook.com/content-restrictions Acesso em 11 maio 2021.

78 O IRIS lançou, em 2020, uma publicação sobre relatórios de transparência que reforça que a divulgação de informações isoladas e descontextualizadas é uma prática comum às plataformas. Ver: KURTZ, Lahis; VIEIRA, Victor. **Obtenção transnacional de conteúdo de comunicações telemáticas na América Latina**: relatório de pesquisa. Instituto de Referência em Internet e Sociedade: Belo Horizonte, 2020. Disponível em: https://irisbh.com.br/wp-content/uploads/2020/04/Obten%C3%A7%C3%A3o-transnacional-de-conte%C3%BAdo-de-comunica%C3%A7%C3%B5es-telem%C3%A1ticas-na-Am%C3%A9rica-Latina.pdf Acesso em: 18 jun. 2021. p. 8

79 Trecho original referenciado: "Transparency can encourage people to be less honest, so increasing deception and reducing reasons for trust: those who know that everything they say or write is to be made public may massage the truth". O'NEILL, Onora. **A question of trust**: The BBC Reith Lectures 2002. Cambridge University Press, 2002. P. 54.

80 O'NEILL, Onora. **A question of trust**: The BBC Reith Lectures 2002. Cambridge University Press, 2002. p. 73.

Crawford[81]. São elencados, no estudo, dez obstáculos para a redução da opacidade: 1. a transparência pode estar desconectada do poder, ou seja, aqueles que serão vistos não são os mesmos capazes de alterar o sistema; 2. a transparência pode inibir diálogo honesto e ameaçar a privacidade; 3. a transparência pode ocultar intencionalmente, pelo excesso de informação ou inserção de informação burocrática/desnecessária em meio à informação importante, desviando atenção ou tornando difícil encontrar o que é relevante; 4. há falsos binários, no sentido de revelar toda ou nenhuma informação, que inviabilizam pensar em tipos diferentes de transparência que acarretam distintas formas de accountability; 5. existe uma ficção de que os indivíduos aos quais a informação é disponibilizada têm, entre si, o mesmo nível de informação. Isso ignora que a informação total, decisões totalmente racionais e participantes totalmente engajados são elementos ficcionais, e não reais, do ambiente de livre troca de ideias; 6. mais transparência não necessariamente constrói mais confiança, pois esta depende da percepção do indivíduo sobre algo, e não sobre a informação provida, necessariamente, sendo que saber mais sobre algo nem sempre torna aquilo mais confiável; 7. a transparência pode ser performativa, com a seleção de informações a serem abertas para revelar apenas o suficiente para controlar o conhecimento sobre o assunto; 8. a transparência pode privilegiar "ver" sobre "compreender", e fornecer informação limitada caso não haja um questionamento sobre a completude da informação e seus limites; 9. existem limitações técnicas e temporais para a transparência: ela é a representação do funcionamento de algo, de forma simplificada, a produção de informação conforme se realiza a atividade não pode ser completa e instantânea; 10. a transparência tem limitações temporais, ou seja, a cultura e o conhecimento que se tem à época limitam o que é possível de ser revelado e conhecido.

Assim, esta pesquisa não parte da ideia de que todo tipo ou toda iniciativa de transparência confere maior poder de fiscalização e responsabilização sobre a atividade de moderação de conteúdo. Por isso, para além de analisar se alguma publicização sobre essas atividades é pautada na regulação, pretende-se avaliar em que medida a regulação confere ao público o potencial de compreender, avaliar e discutir os efeitos da moderação de conteúdo sobre seus direitos. Isto é, o enfoque é na transparência que gera accountability sobre intervenções no conteúdo gerado por usuário.

81 ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New media & society**, v. 20, n. 3, p. 973-989, 2018. Disponível em: https://journals.sagepub.com/doi/10.1177/1461444816676645 Acesso em: 18 jun. 2021.

2. O papel da transparência no cenário regulatório sobre conteúdo online

Como já ressaltado em estudo anterior[82], não se pode esperar que o setor privado seja o único motor na implementação de práticas de transparência – afinal, é também responsabilidade do setor público regular e impor obrigações nesse sentido. Assim, os mecanismos de fiscalização e debate ganham caráter institucional, com a possibilidade de acompanhamento sistemático e ação dirigida para o cumprimento dos requisitos pré-fixados.

Esta seção apresenta os resultados de uma comparação criteriosa dos mecanismos regulatórios estatais sobre transparência na moderação de conteúdo. A seguir, é feita a descrição metodológica e, após, apresenta-se a sumarização de pontos considerados relevantes no escopo do estudo. Por fim, encontram-se os resultados da conferência entre esses parâmetros e a regulação analisada.

2.1. Metodologia de análise

A parte empírica desta pesquisa buscou identificar o estado da arte da regulação sobre transparência na moderação de conteúdo e refletir sobre sua adequação aos direitos humanos e aos princípios que norteiam a internet. Para isso, a análise teórica foi complementada por um estudo comparado de cenários regulatórios, que compreendeu tanto os documentos normativos propriamente ditos (produzidos com a finalidade de orientar expectativas normativas), quanto as decisões judiciais que os aplicam a casos de referência.

Tratando-se de um estudo comparado, esta pesquisa considerou, além do Brasil – local em que ela é desenvolvida –, outros quatro países para compreensão de um cenário transfronteiriço de aplicação de padrões internacionais, obrigações e normas de transparência pelos Estados como estratégia para moderação de conteúdo. Para a seleção dos países foram utilizados os critérios de: 1. variedade continental; 2. disponibilidade da legislação, em fonte oficial, em inglês ou português (idiomas que toda a equipe desta pesquisa possui condições de compreender); 3. existência de um arcabouço jurídico no país, relacionado a direitos e deveres de provedores de conteúdo; 4. movimentações recentes acerca da temática do estudo em diversos setores do país. Assim, a partir da versão mais recente à época do relatório *Freedom on the Net*[83], foram selecionadas as jurisdições: 1. Alemanha, para a Europa, 2. Austrália,

82 RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021

83 FREEDOM ON THE NET. **Map:** internet freedom status**.** Disponível em: https://freedomhouse.

para a Oceania, 3. Brasil, para a América, 4. Índia, para a Ásia, e 5. Quênia, para a África.

Consideraram-se os seguintes fatores para a seleção: no caso da Europa, a Alemanha foi escolhida[84], mesmo não contando com inglês como idioma oficial, devido à influência internacional de sua legislação sobre moderação de conteúdo e governança da internet, além de disponibilizar versões oficiais em inglês de sua legislação relevante à pesquisa. A Austrália foi selecionada pelo histórico de vanguarda na elaboração de legislações de internet com impacto extraterritorial e que ensejaram debates ao redor do globo. Escolheu-se o Brasil na amostra da América, devido aos relevantes debates sobre o assunto que ocorreram em 2020 no Poder Legislativo Federal e ao fato de que os autores desta pesquisa são nacionais deste país. A seleção da Índia deve-se às repercussões internacionais da interferência das plataformas em conteúdo de partidos políticos, além de ser o segundo maior mercado de internet do mundo, o que atrai interesses comerciais das plataformas[85]. O país também aprovou recentemente uma legislação robusta sobre moderação de conteúdo. Por sua vez, o Quênia se destacou e foi selecionado como país para análise devido aos embates realizados na esfera virtual e evidenciados pelo caráter de disputas étnicas e políticas.

É importante realizar uma observação sobre a amostra selecionada e os critérios utilizados: o fato de se ter usado como um critério a escolha de países cuja língua oficial é a inglesa (devido à acessibilidade de material para a equipe) acarretou em seleção de três países que integram a *Commonwealth* (associação de países que se tornaram independentes do Reino Unido, mas permanecem orientados por objetivos sociais e políticos comuns): Austrália, Índia e Quênia. Eventuais pontos e iniciativas em comum entre eles serão analisados tendo também em conta esse contexto – inclusive no que tange ao Guia de Melhores Práticas de Cibersegurança da *Commonwealth*, que dispõe de orientações em matéria de internet[86].

Para compreensão do arcabouço regulatório de cada jurisdição, quatro etapas de pesquisa foram realizadas: 1. pesquisa bibliográfica, a fim de identificar literatura

org/explore-the-map?type=fotn&year=2019 Acesso em: 11 maio 2021.

84 A respeito dessa escolha, foi considerado também o Reino Unido como possível integrante desta amostra, devido ao *Online Security Bill*. Entretanto, como a amostragem compreende um país de cada continente e a regulação da Alemanha contava com repercussão mais robusta em artigos acadêmicos e na mídia internacional, esta foi a escolha para este estudo. Entretanto, considera-se válido realizar novos estudos compreendendo também outras regulações europeias, que são prolíficas nesta matéria, como a do Reino Unido.

85 SORRENTINO, Eduardo. Índia X Twitter: Rede social resiste às mudanças de privacidade. Olhar Digital. 6 jun. 2021. Disponível em: https://olhardigital.com.br/2021/06/06/internet-e-redes-sociais/india-x-twitter-rede-social-resiste-as-mudancas-de-privacidade/. Acesso em 14 jun. 2021.

86 COMMONWEALTH. **Cybersecurity for elections**: a Commonwealth guide on best practices - 2020. Disponível em: https://thecommonwealth.org/sites/default/files/inline/Cybersecurity_for_Elections_PDF_0.pdf Acesso em: 03 fev. 2021.

específica sobre aquela localidade e seu cenário regulatório; 2. levantamento na mídia política, para verificar os dispositivos relevantes, o contexto das propostas e tendências na moderação de conteúdo naquela jurisdição; 3. pesquisa nas bases legislativas de cada país, com a coleta dos documentos regulatórios sobre moderação de conteúdo (leis, diretrizes, projetos de lei, decisões judiciais de referência); 4. consulta a especialistas locais[87], que validaram e revisaram os documentos selecionados para análise regulatória, de forma a contextualizá-los e complementá-los, se necessário. A pesquisa bibliográfica e da mídia política de cada país, bem como a conversa com especialistas, possibilitaram que variações linguísticas, culturais, sociais, políticas e econômicas próprias, que se incluem na legislação de um Estado, fossem consideradas na análise documental.

A análise foi estruturada a partir dos principais requisitos de transparência que poderiam ser abarcados pela regulação. Esses pontos foram levantados pela leitura e seleção de recomendações em relatórios de organizações internacionais, padrões e recomendações da sociedade civil e estudo anteriormente realizado sobre o tema. Essa documentação e os pontos considerados relevantes são apresentados na subseção seguinte e no Apêndice A desta pesquisa.

2.2. Aporte documental para requisitos de transparência

As exigências estatais para transparência na moderação de conteúdo integram um cenário mais amplo, interinstitucional e internacional, de adequação das práticas corporativas aos direitos humanos. Além do pano de fundo teórico apresentado anteriormente, algumas entidades internacionais elaboraram regras e princípios objetivos para possibilitar um cenário de conscientização e debate sobre os limites à expressão online.

Portanto, nesta pesquisa, serão considerados esforços de transparência em sentido amplo – aqueles desenvolvidos no nível teórico e geral – e esforços específicos estabelecidos por instrumentos internacionais. Como esses documentos são considerados norteadores para uma política de regulação de plataformas que respeite os direitos humanos, eles também podem servir de parâmetro para avaliar as políticas normativas adotadas. Assim, destacam-se alguns instrumentos e, em especial, alguns pontos cujo foco é a abertura das práticas ao conhecimento (e questionamento) do público.

87 Nossos agradecimentos a: Akriti Bopanna, Gerente de Policy e Advocacy na Internet Society, e Rohin Garg e Apar Gupta, da Internet Freedom Foundation, pelas contribuições na pesquisa sobre a Índia; Daphnee Prates Iglesias, experta em digitalização e políticas públicas; German Chancellor Fellow da Fundação Alexander von Humboldt (Alemanha), pelas contribuições na pesquisa sobre a Alemanha ; Muthoni Muriithi, Gerente Sênior em Políticas de Gênero na World Wide Web Foundation, pelas contribuições na pesquisa sobre o Quênia.

A *Electronic Frontier Foundation* ocupa papel de destaque nesse contexto, ao participar ativamente da coordenação de dois dos principais instrumentos mobilizados pela sociedade civil. Os Princípios de Manila[88] e os Princípios de Santa Clara[89] são voltados a dois dos pontos mais cruciais da moderação de conteúdo online: 1. a responsabilidade de intermediários em relação à atividade de moderação; e 2. a transparência sobre essas práticas. Ou seja, as preocupações estão em estabelecer quais seriam as boas práticas procedimentais e em manter sob amplo escrutínio público as decisões tomadas.

Conforme mencionado, a transparência inicialmente foi concebida como parte e requisito para a legitimidade governamental e, com as mudanças e demandas do desenvolvimento da sociedade, passou a ser praticada e exigida por entes não-estatais. A observância do devido processo, embora de origem e essencialidade estatal, é destacada nos Princípios de Manila como esforço a ser adotado pelas plataformas – entes privados, em sua maioria – devido à importância de se garantir legitimidade ao processo de moderação de conteúdo.

Assim, devido processo e transparência dialogam por atualmente alcançarem entes não estatais, devido à reconfiguração social, e por garantirem proteção às pressões arbitrárias feitas sobre plataformas para restringirem conteúdos. Em diversos momentos, o devido processo aumenta as exigências de transparência sob a atuação das plataformas e, em outros, as exigências de transparência garantem o acompanhamento das garantias previamente definidas para a restrição de conteúdo.

Ainda no âmbito da sociedade civil organizada, a *Access Now* publicou um documento[90] com recomendações para governança de conteúdo, que inclui diversas dimensões e preocupações. Estas vão desde boas práticas de moderação por plataformas e governos, discutindo os diferentes formatos de regulação, até medidas de transparência e construção participativa de regulação sobre conteúdo online.

A *Global Network Initiative*, por sua vez, é uma rede de empresas de tecnologia da informação, sociedade civil, acadêmicos e investidores que reúne preocupações do setor privado com uma atuação alinhada a direitos humanos face à atuação governamental. Sua contribuição de referência são os Princípios GNI de Liberdade de

88 EFF - Electronic Frontier Foundation et al. **Princípios de Manila sobre responsabilidade civil de intermediários**. Disponível em: https://manilaprinciples.org/pt-br/principles.html Acesso em: 14 jun. 2021.

89 EFF - Electronic Frontier Foundation et al. **Santa Clara Principles on transparency and accountability in content moderation**. Disponível em: https://santaclaraprinciples.org/ Acesso em: 14 jun. 2021.

90 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones sobre gobernanza de contenido**: una guía para legisladores, reguladores y encargados de políticas empresariales. Access Now, 2020. Disponível em: https://www.accessnow.org/cms/assets/uploads/2020/03/Recomendaciones-Gov-Contenidos.pdf Acesso em: 18 jun. 2021.

Expressão e Privacidade[91]. Eles são voltados, sobretudo, a solicitações governamentais de moderação de conteúdo e avaliam padrões e medidas que podem ser adotadas pelas companhias.

A Organização das Nações Unidas para Organização Educacional, Científica e Cultural (UNESCO), em seu relatório *Fostering Freedom Online*[92], avaliou o papel de intermediários online. Nesse documento, que reconhece a *Global Network Initiative* como uma importante referência, são examinados os riscos da restrição de conteúdo online, os tipos de regulação sobre responsabilidade de intermediários e as lacunas nas práticas de alguns provedores de serviços na internet em relação à transparência das regras, notificação do usuário e mecanismos de recurso.

Sob o prisma da liberdade de expressão, a Organização das Nações Unidas preocupa-se com boas práticas em diversos setores e também com atitudes a serem evitadas no âmbito governamental – como a responsabilização de intermediários por remover conteúdo ilícito, ou a determinação de um prazo para tanto, por exemplo. Os relatórios elaborados por David Kaye, na condição de Relator Especial para a Liberdade de Expressão, apontam para a necessidade de leis focadas na transparência e na amplificação do poder de decisão do público sobre suas ações online[93]. Algumas recomendações propõem evitar que terceiros alheios ao controle social tenham poder para determinar o que é ou não conteúdo legítimo[94].

Uma referência para as demais iniciativas é o documento de Princípios Orientadores sobre Empresas e Direitos Humanos[95], chamados Princípios Ruggie (em menção ao Representante Especial do Secretário-Geral das Nações Unidas, Professor John Ruggie). No princípio nº 21, é expressa a necessidade de a empresa informar as medidas que toma para mitigar os riscos em qualquer interferência em direitos humanos (como é o caso da livre expressão, do acesso à informação ou de não-discriminação), o que permite aos afetados por sua atividade tomar conhecimento e demandar respostas sobre questões emergentes.

---

91 GNI - Global Network Initiative. **The GNI Principles Principles on Freedom of Expression and Privacy**. Disponível em: https://globalnetworkinitiative.org/gni-principles/ Acesso em: 14 jun. 2021.

92 MACKINNON, Rebecca; HICKOK, Elonnai; BAR, Allon; LIM, Hae-In. **Fostering freedom online:** the role of internet intermediaries. UNESCO series on internet freedom. França, 2014. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000231162_eng Acesso em: 14 jun. 2021.

93 ONU - Organização das Nações Unidas. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. **Report on Content Regulation**. Symbol A/HRC/38/35. 6 abr. 2018. Disponível em: https://www.undocs.org/A/HRC/38/35 Acesso em: 14 jun. 2021. p. 16

94 Ibidem, p. 20

95 ONU - Organização das Nações Unidas. **Princípios Orientadores sobre Empresas e Direitos Humanos**. Implementando os Parâmetros "proteger, respeitar e reparar" das Nações Unidas. 16 jun. 2011. Trad. Ministério das Relações Exteriores. Disponível em: https://www.gov.br/mdh/pt-br/assuntos/noticias/2019/outubro/Cartilha_versoimpresso.pdf Acesso em 11 maio 2021.

Os princípios Ruggie embasam o Guia dos Direitos Humanos para Utilizadores de Internet[96], do Conselho da Europa, o qual também promove a necessidade da moderação de conteúdo a fim de evitar a propagação de conteúdo discriminatório, por exemplo. Entretanto, ressalta que essa ingerência na liberdade de expressão deve ocorrer dentro de certos limites, que exigem transparência sobre a filtragem e a desindexação de conteúdos[97]. Tais limites são estabelecidos com base em um documento de Recomendações dos Ministros dos Estados-Membros sobre a proteção de direitos humanos em serviços de redes sociais[98].

Às questões e documentos de referência já mencionados, agregam-se também pontos de um estudo realizado pela equipe do IRIS que consistiu na análise da transparência proporcionada pelas políticas de comunidade. Nele, foram identificadas lacunas em temas relevantes, sobretudo na publicização de critérios de detecção e análise do conteúdo moderado. Ou seja, há opacidade sobre as instruções seguidas pelos moderadores ou inseridas no código de programas de tratamento automatizado, que delineiam esta atuação[99]. Esse fator tem reflexos diferentes em cada caso.

Quando realizada por equipes humanas, espera-se que a moderação tenha maior potencial de análise de contexto e interpretação, porém as condições de treinamento e de trabalho afetam diretamente as práticas adotadas. Estudos também demonstram a ausência de um ambiente adequado de trabalho para essas pessoas, o que viola seus direitos[100]. Assim, maior transparência sobre quais são as condições de contratação, treinamento e realização do trabalho também é relevante para discussões envolvendo a qualidade da moderação realizada. No contexto da moderação por algoritmos, existe risco de remoção de conteúdo legítimo devido a leituras equivocadas pelo menor potencial adaptativo da máquina, que não é afetada pelo seu entorno físico, mas pelas informações com as quais é programada e alimentada. Nesse sentido, a transparência envolveria saber que testes foram efetuados para prevenir discriminação e remoção de conteúdo legítimo, o que possibilitaria a sugestão de testes complementares e ajustes.

96 CE - Conselho da Europa. **Guia dos Direitos Humanos para os Utilizadores de Internet.** 16 abr. 2014. Disponível em: http://www.odionao.com.pt/media/5185/GuiaDireitosHumanosUtilizadoresInternet.pdf Acesso em 11 maio 2021.

97 Ibidem, p. 35-38

98 CE - Conselho da Europa. **Recommendation CM/Rec(2012)4** of the Committee of Ministers to member States on the protection of human rights with regard to social networking services. 4 abr. 2012. Disponível em: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b Acesso em: 11 maio 2021.

99 RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021. p. 68, 76.

100 ROBERTS, Sarah T.. **Behind the Screen**: Content Moderation in the Shadows of Social Media. New Haven: Yale University Press, 2019

Os principais pontos contidos nesses documentos que podem nortear regulações acerca da moderação de conteúdo, com enfoque na transparência dessas atividades, foram sistematizados para a elaboração de um formulário de análise, conforme o Apêndice A. Feita a exposição dos documentos que embasaram os critérios de análise, passa-se à apresentação dos resultados quanto à observância desses parâmetros nos distintos cenários regulatórios nacionais selecionados.

2.3. Apresentação dos resultados

Submeteram-se as notícias e os documentos regulatórios – listados no Apêndice B – da amostra de cinco países supramencionada a uma análise estruturada com base nos tópicos destacados na seção anterior. Para cada país, preencheu-se um formulário (Apêndice C) , com referência aos dispositivos regulatórios e fontes pertinentes.

Esta seção apresenta uma síntese dos resultados, tanto agrupados por país quanto o comparativo entre eles, com pontos de semelhança e diferenças entre cada contexto regulatório. Foi possível perceber quais os enfoques regulatórios e contextos onde o conteúdo online é abordado e identificar lacunas e elementos a serem aprimorados em todas as jurisdições. Para cada uma, são relatados: 1. Os eventos midiáticos e políticos relevantes para a conformação do cenário regulatório sobre moderação de conteúdo online; 2. o conteúdo da regulação analisada e como ela se integra ao cenário jurídico desse país; 3. elementos sobre transparência que a regulação aborda ao tratar sobre a atividade das plataformas de conteúdo; 4. pontos sobre moderação de conteúdo que são abordados na regulação e sua correlação com transparência.

2.3.1. Alemanha

a) Contexto político e midiático

A Alemanha é internacionalmente reconhecida pela ênfase que a legislação nacional atribui à defesa dos direitos autorais. Os enunciados alemães sobre propriedade intelectual são notoriamente rígidos e compõem um dos arcabouços legais mais extensos sobre a temática.

Um exemplo icônico da severidade dos enunciados no país são as multas aplicadas àqueles que violam a lei. Na Alemanha, observa-se uma grande ênfase no combate a serviços de *streaming* ilegais – especificamente, os que não detém os direitos de reprodução das obras disponibilizadas[101] –, e o uso dessas aplicações pode acarretar

101 Um estudo publicado em 2018 revela que, após a queda do principal site para *streaming* não licenciado de vídeos na Alemanha – o kino.to –, em 2011, diversos serviços similares – novos ou já previamente existentes – foram utilizados para suprir a demanda por conteúdo dessa natureza.

em responsabilização civil na forma de multas com valores elevados – em torno de € 1.000, em média[102]. O mesmo pode ser dito com relação ao download de materiais protegidos por direitos autorais através de serviços de torrent ou outras formas de compartilhamento *peer-to-peer*.

Além disso, até outubro de 2017, a possibilidade de responsabilização e multa podia ser estendida ao dono da rede de internet utilizada, mesmo que a infração fosse cometida por um visitante ou um desconhecido. Isso se dava em decorrência do chamado *Störerhaftung*[103]: um mecanismo legal que atraía para o proprietário de uma rede a responsabilidade por qualquer ato ilícito realizado através dessa conexão[104]. Em decorrência disso, até a supressão desse enunciado por uma alteração da Lei de Telecomunicações – *Telekommunikationsgesetz*[105] –, redes Wi-Fi disponíveis a terceiros (em estabelecimentos comerciais, por exemplo) eram escassas no país.

Seguindo essa tendência regulatória de grande proteção aos direitos de propriedade intelectual, em 2012 a proposta de Direitos Autorais Auxiliares Para Editores de Imprensa (*Leistungsschutzrecht für Presseverleger*) – aprovada na forma das Seções 87f e subsequentes da Lei de Direitos Autorais (*Urheberrechtsgesetz*[106]) – buscou garantir de forma mais extensiva os direitos autorais de plataformas de veiculação de notícias e incluiu previsões específicas para o meio digital. Essa legislação foi aprovada no Parlamento alemão em março de 2013, em meio a grande controvérsia[107] sobre seu

O estudo aponta para o fato de que há uma grande facilidade de migração dos usuários para sites correlatos quando um deles é tornado indisponível no país, o que resultou em uma queda não muito significativa dos índices de acesso ao streaming não licenciado. AGUIAR, L., CLAUSSEN, J., PEUKERT, C. . Catch Me if You Can: Effectiveness and Consequences of Online Copyright Enforcement. **Information Systems Research**, 29(3), p. 656–678. 2018. Disponível em: https://research-api.cbs.dk/ws/portalfiles/portal/57343396/j_rg_claussen_et_al_catch_me_if_you_can_acceptedversion.pdf. Acesso em: 19 maio 2021.

102 European Consumer Centre Germany. **Illegal downloads in Germany.** Disponível em: https://www.evz.de/en/shopping-internet/illegal-downloads-in-germany.html. Acesso em: 19 maio 2021.

103 Em tradução literal, esse termo significa algo como "responsabilidade por falha". Trata-se de um mecanismo que, aplicado ao meio digital, estendia a responsabilidade por ilícitos digitais também aos chamados "facilitadores" ou "possibilitadores" daquele ilícito. Trata-se de um mecanismo derivado do §1004 do Código Civil alemão (*Bürgerliches Gesetzbuch*), e utilizado para diversas áreas correlatas do Direito.

104 DEUTSCHE WELLE. Germany to abolish provider liability law, open path to more free WiFi. **Deutsche Welle**. Disponível em: https://www.dw.com/en/germany-to-abolish-provider-liability-law-open-path-to-more-free-wifi/a-19249024. Acesso em: 19 maio 2021.

105 ALEMANHA. BUNDESTAG. **Lei de Telecomunicações**. Berlim, 22 de junho de 2004. Disponível em: https://germanlawarchive.iuscomp.org/?p=692. Acesso em 17 jun. 2021.

106 ALEMANHA. BUNDESTAG. **Lei de Direitos Autorais e Direitos Correlatos**. Berlim, 09 de setembro de 1965. Disponível em: https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html. Acesso em 17 jun. 2021.

107 Os dispositivos em questão causaram grande polarização argumentativa. Pôde-se observar, por exemplo, indivíduos que entendiam pela legitimidade desses enunciados, defendendo que a reprodução de trechos de conteúdo protegido por direitos autorais para fins econômicos – como é o caso das plataformas de aglomeração de notícias como o Google Notícias, que lucram com essa prática

conteúdo. O principal ponto de atrito da normativa foram os dispositivos sobre a estipulação de taxas monetárias, diretamente com os detentores de direitos autorais por publicações jornalísticas, para que plataformas pudessem incluir trechos (*snippets*) dessas obras junto aos links para as notícias[108].

Para além dos enunciados relativos a direitos autorais no ambiente regulatório alemão, medidas legislativas foram tomadas para combater um aumento na disseminação de discurso de ódio na rede. Essa situação se intensificou no país com a alta nos índices de imigração para a Alemanha – principalmente de refugiados provenientes da Síria. Isso, por sua vez, culminou na criação da Lei de Regulação da Internet (*Netzdurchsetzunggesetz*, popularmente conhecida como "NetzDG"[109]) em 2017, após diversas tentativas malsucedidas de implementação de um sistema de autorregulação pelas plataformas digitais[110].

b) Conteúdo da regulação

O cenário regulatório-jurisprudencial alemão relativo ao compartilhamento de conteúdo por usuários de plataformas digitais tem, já há alguns anos, um enfoque na proteção de direitos autorais. A Lei de Direitos Autorais do país é um fundamento recorrente para remoção de conteúdo online. Apesar disso, ela não estabelece enunciados específicos sobre moderação de conteúdo digital. Ela motiva essa prática porque prevê que pessoas detentoras de direitos autorais possuem o direito exclusivo sobre suas obras protegidas por propriedade intelectual.

– justificava a cobrança de royalties. Por outro lado, a parte das mobilizações que se posicionou em desfavor das medidas defendia que a proibição da veiculação de trechos introdutórios – "manchetes" desses conteúdos – poderia levar a desinformação decorrente da leitura e interpretação insuficiente apenas dos títulos desses materiais jornalísticos. Também foi alegado que a veiculação desses trechos do material serviria como forma de angariar interesse dos leitores e, consequentemente, aumentar gratuitamente a quantidade de acessos de usuários aos portais de veiculação de notícias. Uma análise sumarizando as principais críticas à legislação pode ser encontrada em: BITKOM. **Ancillary Copyright for Publishers**: taking stock in germany. Berlim: Bitkom, 2015. Disponível em: https://www.bitkom.org/sites/default/files/file/import/160406-Leistungsschutzrecht-EN1.pdf. Acesso em: 19 maio 2021.

108 A Lei de Direitos Autorais Auxiliares Para Editores de Imprensa foi posteriormente utilizada como referência para a criação de dispositivos similares na Diretiva da União Europeia Sobre Direitos de Autor. Essa diretiva, que entrou em vigor em junho de 2019 e passará a surtir efeitos após 2 anos de sua vigência, também motivou grande controvérsia no cenário europeu e internacional. As previsões relativas a direitos autorais para a imprensa são em grande parte similares às da legislação alemã, e representaram um dos principais pontos de discordância que permearam o percurso legislativo da Diretiva. Para mais informações, conferir: FERRI, F. The dark side(s) of the EU Directive on copyright and related rights in the Digital Single Market. **China-EU Law Journal** 2020. Disponível em: https://doi.org/10.1007/s12689-020-00089-5. Acesso em: 19 maio 2021.

109 ALEMANHA. BUNDESTAG. **Lei de Regulação da Internet**. Berlim, 01 de setembro de 2017. Disponível em: https://germanlawarchive.iuscomp.org/?p=1245. Acesso em 17 jun. 2021.

110 HELDT, Amélie Pia. Reading between the lines and the numbers: an analysis of the first NetzDG reports. **Internet Policy Review**, v. 8, n. 2, 2019. Disponível em: https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports Acesso em: 18 jun. 2021.

De forma similar, o Código Criminal alemão (*Strafgesetzbuch*[111]) não apresenta previsões específicas sobre conteúdo digital e métodos para controle e moderação de conteúdos considerados danosos. A lei, contudo, criminaliza a disseminação de discurso de ódio em qualquer meio e é empregada também para fundamentar decisões de responsabilização daqueles que compartilham conteúdo online dessa natureza.

A NetzDG, que entrou em vigor no ano de 2017, gerou controvérsias[112] globais a respeito de seu modelo de remoção de conteúdo. A regulação foi criada com a finalidade de desincentivar a disseminação de conteúdo danoso e desinformativo online. O modelo proposto previu alterações no regime de responsabilidade das plataformas digitais com pelo menos 2 milhões de usuários na Alemanha.

Finalmente, o Tratado Estatal de Mídia (*Medienstaatsvertrag* – MStV[113]) entrou em vigor em novembro de 2020 e atualizou os enunciados do antigo Tratado Interestadual de Transmissão (*Rundfunkstaatsvertrag* – RStV[114]). A regulação reconhece a existência de novas tecnologias de transmissão de conteúdo, a exemplo da internet, e define novas obrigações para as empresas, incluindo emissoras tradicionais, empresas análogas a emissoras tradicionais, plataformas de agregação de mídia gerada por terceiros (intermediários de mídia), provedores de interface do usuário, redes sociais, lojas digitais etc. Entre esses enunciados, cabe apontar aqueles que dizem respeito à obrigação de transparência algorítmica por parte de emissoras, empresas análogas a emissoras, interface do usuário e intermediários de mídia[115]. Os novos enunciados obrigam os provedores desses serviços a fornecer informações nítidas e facilmente acessíveis – inclusive aos usuários – sobre o funcionamento de seus algoritmos, incluindo a forma como o conteúdo é ranqueado e selecionado para visualização.

111 ALEMANHA. BUNDESTAG. **Código Criminal**. Berlim, 15 de maio de 1871. Disponível em: https://www.gesetze-im-internet.de/englisch_stgb/. Acesso em 17 jun. 2021.

112 TWOREK, Heidi; LEERSSEN, Paddy. **An Analysis of Germany's NetzDG Law**. Abr. 2019. p. 4-8. Disponível em: https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf. Acesso em: 19 maio 2021.

113 ALEMANHA. BUNDESLÄNDER. **Tratado de Mídia Estatal**. Berlim, 28 de abril de 2020. Disponível em: https://www.gesetze-bayern.de/Content/Document/MStV/true. Acesso em 17 jun. 2021.

114 ALEMANHA. BUNDESLÄNDER. **Tratado Interestadual de Transmissão**. Berlim, 31 de agosto de 1991. Disponível em: https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=2&gld_nr=2&ugl_nr=2251&bes_id=12784&aufgehoben=N&menu=1&sg=0. Acesso em 17 jun. 2021.

115 ALGORITHM WATCH. **Germany's new media treaty demands that platforms explain algorithms and stop discriminating. Can it deliver?** Disponível em: https://algorithmwatch.org/en/new-media-treaty-germany/. Acesso em 19 maio 2021.

c) Requisitos de transparência

- **Quanto às obrigações das plataformas**
  Há determinações para que as plataformas digitais que recebem mais de 100 denúncias sobre conteúdo ilegal a cada ano elaborem relatórios de transparência semestralmente[116]. Esses relatórios devem explicitar a forma como essas denúncias foram endereçadas e incluir uma série de informações sobre elas, que serão enumeradas no tópico abaixo sobre requisitos para relatórios de transparência[117]. Todas as decisões das plataformas relativas a denúncias sobre ilegalidade de conteúdo devem ser informadas e justificadas tanto para a pessoa que publicou o material quanto para a que o denunciou.[118]

  Além disso, o Tratado Estatal de Mídia[119] enuncia que plataformas em geral – desde plataformas de mídia e interfaces de usuário até as consideradas apenas intermediárias de mídia – devem disponibilizar publicamente informações sobre a forma como seus algoritmos realizam a seleção e ranqueamento do conteúdo que é apresentado ao público.

- **Quanto à restrição de conteúdo ilícito ou com base em políticas de comunidade**
  Não há previsões

- **Quanto a políticas de comunidade**
  Não há previsões

- **Quanto a relatórios de transparência**
  O NetzDG estabelece que os relatórios de transparência derivados dessa lei devem conter o número de denúncias que resultaram em remoção ou bloqueio de conteúdo pela plataforma, subdivididas de acordo com o motivo para essas denúncias e a sua origem (se foram realizadas por usuários ou por entidades representativas). Adicionalmente, os relatórios devem incluir: dados gerais sobre os esforços empregados pela plataforma para combater atividades ilegais; o número de denúncias realizadas, subcategorizadas pelo tipo de autor, pelo motivo da denúncia, entre outros; o tempo de resposta às denúncias; a existência de comunicação ao autor da denúncia e ao autor do conteúdo denunciado sobre a decisão tomada pela plataforma; entre outros.[120]

  Apesar das discussões causadas pelo NetzDG, observa-se no cenário alemão

116 Lei de Regulação da Internet, Seção 2 (1).
117 Lei de Regulação da Internet, Seção 2 (2).
118 Lei de Regulação da Internet, Seção 3 (2).
119 Tratado Estatal de Mídia. Subseção 2, §85, e Subseção 3, §93.
120 Lei de Regulação da Internet, Seção 2 (2).

uma preocupação considerável com os mecanismos de transparência relativos à moderação de conteúdo online. Uma alteração a essa normativa[121], proposta em abril de 2020 e ainda em tramitação no momento da redação do presente estudo, prevê mais direitos para os usuários das plataformas digitais por meio da implementação de mais obrigações de transparência em decisões que levem essas empresas a remover conteúdo compartilhado.

Mais especificamente, em primeiro lugar, a referida modificação enuncia o direito dos usuários a recorrerem de decisões das plataformas sobre denúncias de conteúdo danoso. Pretende, assim, ampliar o direito ao contraditório e reduzir a incidência de remoções de conteúdo errôneas e bloqueios de contas equivocados. Essa alteração imputa aos relatórios de transparência previstos pelo NetzDG[122] a obrigação de informar aos usuários sobre os mecanismos de transmissão dessas reclamações, os critérios envolvidos na moderação do conteúdo em questão e a forma como o processo de exame do conteúdo foi realizado.

Além disso, a mudança busca facilitar o acesso das pessoas às ferramentas para denúncia de conteúdo, tornando seu uso mais compreensível, a fim de encorajar a comunicação de ilegalidade desses materiais quando necessário (Seção 3(1)). Por fim, a normativa busca reforçar a necessidade de transparência significativa por parte das plataformas e altera a Lei de Telecomunicações para facilitar que autoridades judiciárias obtenham das plataformas informações sobre perfis identificados como veiculadores de conteúdo danoso.

d) Obrigações quanto à moderação de conteúdo

As plataformas têm o ônus de monitorar as denúncias de usuários nas postagens realizadas em seus serviços.[123] Mais especificamente, denúncias sobre conteúdo danoso devem ser avaliadas em até 24 horas para remoção de conteúdo "obviamente ilícito" e em até 7 dias para remoção ou bloqueio de conteúdo ilícito, após a avaliação das informações veiculadas.

Grande parte das críticas concentra-se em torno do receio de que as medidas de remoção de conteúdo representam um precedente perigoso para a moderação de conteúdo online. Isto porque, conforme alegam os opositores[124] do NetzDG, a lei

121 Trata-se do *Entwurf eines Gesetzes zur Änderung des Netzwerkdurchsetzungsgesetzes*, traduzido livremente para "Projeto de lei para alteração do NetzDG".

122 Lei de Regulação da Internet. Seção 2 (2).

123 Lei de Regulação da Internet. Seção 3 (2).

124 Há diversas entidades entre os opositores da referida legislação. O Relator Especial das Nações Unidas para a promoção e proteção dos direitos à liberdade de opinião e expressão, David Kaye, apontou que a regulação proposta resultaria em obrigações excessivas para as plataformas digitais, e que a lei é incompatível com declarações internacionais de direitos humanos, como o Pacto Internacional de Direitos Civis e Políticos. Ver: KAYE. David. **Mandate of the Special Rapporteur on the promotion**

não apenas estabelece às plataformas um dever de fiscalização que usualmente é atribuído ao Estado, como também abre portas para o uso de filtros de conteúdo que podem levar à restrição indevida de conteúdo. Consequentemente, pode levar a danos para a liberdade de expressão no país. Além disso, o volume de conteúdo publicado e denunciado nas plataformas tornaria o uso de filtros a única maneira viável para que essas determinações fossem obedecidas pelas plataformas sujeitas aos enunciados legais. As remoções indevidas de conteúdo dos usuários, por sua vez, poderiam institucionalizar um cenário de censura digital na internet alemã.

Ainda com relação ao NetzDG, em março de 2021 entrou em vigor uma nova adição: trata-se da Lei para Combate ao Extremismo de Direita e ao Crime de Ódio (*Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität*[125]). A legislação inseriu ao NetzDG e à Lei de Telecomunicações alemã enunciados que obrigam as plataformas digitais a reportarem para as entidades governamentais informações sobre o compartilhamento de conteúdo extremista de direita[126] e discurso de ódio. Informações sobre os usuários que compartilharam esses conteúdos também devem ser prestadas ao Estado.

As alterações ao NetzDG – tanto a supracitada quanto a relativo ao direito de contestação de decisões sobre moderação de conteúdo – também têm sido alvo de consideráveis críticas. Primeiramente, observa-se que as preocupações relativas à atribuição de deveres fiscalizatórios usualmente estatais para as plataformas digitais não foram endereçadas: na verdade, os acréscimos ao texto legal ampliam as responsabilidades desses entes privados, o que intensifica o cenário regulatório anteriormente criticado. Além disso, a adição relativa ao extremismo político e ao discurso de ódio prevê a aplicação de multas para as plataformas que falharem em reportar as atividades criminosas para as autoridades públicas. Ao mesmo tempo, não há repercussões previstas para as plataformas que equivocadamente reportarem conteúdo lícito. Dessa forma, surge também o receio de que sejam

and protection of the right to freedom of opinion and expression. 01 de junho de 2017. Disponível em: https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf. Acesso em: 17 jun. 2021. Nesse mesmo sentido, entidades vinculadas a diversos setores – notoriamente, a mídia de veiculação de notícias – uniu-se à Global Network Initiative para manifestar receio quanto às implicações práticas do NetzDG. Ver: GLOBAL NETWORK INITIATIVE. **Proposed German Legislation Threatens Free Expression Around the World**. 20 de abril de 2017. Disponível em: https://globalnetworkinitiative.org/proposed-german-legislation-threatens-free-expression-around-the-world/. Acesso em: 17 jun. 2021.

125 ALEMANHA. BUNDESTAG. **Lei para Combate ao Extremismo de Direita e ao Crime de Ódio**. Berlim, 30 de março de 2021. Disponível em: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/Bgbl_Bekaempfung_Rechtsextremismus_Hasskriminalitaet.pdf?__blob=publicationFile&v=3. Acesso em 17 jun. 2021.

126 Chamou a atenção da equipe a escolha por um mecanismo tão específico na legislação alemã. A redação específica para combate a um polo do espectro político pode estar relacionada aos receios advindos do passado nazista do país, marcado pelo holocausto. Ainda assim, importa pontuar que a adoção de parâmetros legais demasiadamente restritivos para moderação de conteúdo é contraindicada pelas referências internacionais para moderação apontadas por este estudo.

reportadas em massa publicações não danosas. Isso potencialmente submeteria as pessoas a investigações estatais de forma desnecessária e intrusiva.

2.3.2. Austrália

a) Contexto político e midiático

Já há alguns anos, a Austrália protagoniza discussões relativas a instrumentos normativos para regulação do meio digital. Diversas legislações australianas angariam grande atenção midiática, muitas vezes acompanhada de críticas por parte da sociedade civil e da comunidade técnico-científica.

Um dos pontos de controvérsia recorrentes apontados pelos críticos das normativas do país trata da falta de abertura das autoridades australianas para consultas públicas e reuniões técnicas com profissionais das áreas afetadas pelas leis. Também nesse sentido, críticas são repetidamente dirigidas ao ritmo acelerado do trâmite legislativo, sem que haja tempo suficiente para o debate e o refinamento necessários para projetos com tantas repercussões práticas.

Dentre as normativas analisadas, a primeira que passou por controvérsias dessa natureza foi a Lei para Emenda de Legislação Relativa a Telecomunicações e Outros (Assistência e Acesso) – *Telecommunications and Other Legislation Amendment (Assistance and Access) Act*[127], em 2018. Trata-se de uma legislação que versa sobre técnicas criptográficas e que foi aprovada às pressas com o objetivo de garantir, com o máximo de celeridade, a defesa nacional frente a ameaças de ataques terroristas[128]. Contudo, a rapidez e a unilateralidade do processo legislativo, somadas às previsões legais consideradas demasiadamente severas por críticos internacionais, resultaram no apelido "lei australiana anticriptografia".

Debate similar quanto à unilateralidade do processo legislativo australiano ocorreu no ano de 2019, durante a tramitação da Emenda ao Código Penal (Compartilhamento de Material Abominavelmente Violento) – *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act*[129]. Essa modificação ao texto legal foi motivada pela transmissão ao vivo do atentado de Christchurch, na Nova Zelândia, que recebeu grande atenção midiática e regulatória. O atentado armado a uma mesquita na

127 AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei para Emenda de Legislação Relativa a Telecomunicações e Outros (Assistência e Acesso)**. Camberra, 08 de dezembro de 2008. Disponível em: https://www.legislation.gov.au/Details/C2018A00148. Acesso em: 17 jun. 2021.

128 Mais especificamente, a tramitação da lei foi acelerada em decorrência de um acordo legislativo, por meio do qual o *Labor Party* – partido de oposição política do país na época – concordou em aprovar a lei sem deliberação relativa às 173 emendas propostas ao texto legal. Isso se deu em virtude do atentado terrorista do Distrito Central de Negócios de Melbourne, que sensibilizou a população do país e culminou no senso de urgência pela aprovação de leis preventivas de eventos dessa natureza.

129 AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Emenda ao Código Penal** (Compartilhamento de Material Abominavelmente Violento). Camberra, 05 de abril de 2019. Disponível em: https://www.legislation.gov.au/Details/C2019A00038. Acesso em: 17 jun. 2021.

cidade de Christchurch culminou na morte de cerca de 50 frequentadores do templo. O episódio foi transmitido ao vivo pelo próprio atirador por meio da plataforma de *streaming* do Facebook, e permaneceu acessível por um período considerável após seu *upload*. Por isso, motivou discussões regulatórias sobre práticas de moderação de conteúdo online[130].

O evento culminou na *Christchurch Call to Eliminate Terrorist & Violent Extremist Content Online*, um movimento composto por governos diversos e empresas de tecnologia, com o objetivo de combater o terrorismo e o conteúdo extremista e violento no meio digital. A emenda foi proposta no Parlamento e aprovada no Senado no mesmo dia, sem abertura para participação pública nos debates. Relatores Especiais da *United Nations Human Rights Special Procedures*, por exemplo, alegaram que gostariam de manifestar suas preocupações quanto a aspectos controversos do conteúdo legal, mas foram impedidos de o fazer em decorrência da pressa excessiva do Legislativo australiano.

O mais recente atrito legislativo data de 2020 e estendeu-se até o início de 2021: trata-se da Lei de Emenda às Leis do Tesouro (Código de Negociação Obrigatório para Plataformas Digitais e Mídia de Notícias) – *Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Bill 2020*[131]. Entre outras disposições, a lei prevê um código de negociação obrigatório para que plataformas paguem *royalties* aos veículos de mídia jornalística pela agregação de material dessa natureza distribuído dentro de seus ambientes digitais. A versão inicial da lei foi submetida ao Legislativo australiano após um estudo[132] realizado pela Comissão Australiana de Competição e Consumidores[133] (ACCC). Também foi realizada uma consulta pública[134] para que representantes de setores da sociedade interessados na matéria pudessem colaborar para as recomendações da ACCC ao projeto de lei.

A lei inicialmente aprovada gerou resistência por parte de plataformas digitais como Facebook e Google, que alegaram inviabilidade de seus modelos de negócios perante às determinações legais. Embora a Google, apesar das críticas, tenha demonstrado interesse em dialogar com as autoridades australianas para alcançar

130 Para mais informações, acessar: **Christchurch Call**: to eliminate terrorist & violent extremist content online. to Eliminate Terrorist & Violent Extremist Content Online. Disponível em: https://www.christchurchcall.com/index.html. Acesso em: 19 maio 2021.

131 AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei de Emenda às Leis do Tesouro** (Código de Negociação Obrigatório para Plataformas Digitais e Mídia de Notícias). Camberra, 09 de dezembro de 2020. Disponível em: https://www.legislation.gov.au/Details/C2020B00190. Acesso em: 17 jun. 2021.

132 AUSTRALIAN COMPETITION & CONSUMER COMMISSION. **Digital Platforms Inquiry:** Final Report. Junho de 2019. Disponível em: https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf. Acesso em: 22 jun. 2021.

133 Tradução livre de "Australian Competition & Consumer Commission".

134 AUSTRALIAN COMPETITION & CONSUMER COMMISSION. **News Media Bargaining Code:** Draft Legislation. 31 de julho de 2020. Disponível em: https://www.accc.gov.au/focus-areas/digital-platforms/news-media-bargaining-code/draft-legislation. Acesso em: 21 jun. 2021.

uma regulação considerada viável pela empresa[135], o Facebook adotou medidas drásticas para protestar contra a legislação adotada. A empresa chegou até mesmo a cessar parcialmente suas atividades no país, em resposta ao texto inicial da lei. Mais especificamente, o Facebook restringiu totalmente o compartilhamento de material jornalístico em sua plataforma, até que melhores condições de operação fossem garantidas pelo Legislativo[136]. Em decorrência da pressão política, a legislação foi alterada em fevereiro de 2021. Com as mudanças, observou-se um retrocesso nas críticas e ameaças por parte das grandes plataformas digitais – em especial, por parte do Facebook, que voltou atrás quanto à restrição de compartilhamento de notícias.

b) Conteúdo da regulação

No cenário australiano, observou-se uma prevalência da regulação por meio de legislações, algumas das quais já foram mencionadas acima e são detalhadas a seguir. Foi analisado também o Projeto de Lei para Segurança Online (*Online Safety Bill*[137]), que reproduz alguns dos enunciados da Lei Para Melhorar a Segurança Online (*Enhancing Online Safety Act*[138]) e acrescenta algumas previsões ao seu texto legal. A legislação australiana analisada é marcada por enunciados que sugerem uma ampliação da responsabilidade de plataformas digitais por conteúdo gerado por usuários. Nesse sentido, as normativas preveem, em certa escala, o dever de autofiscalização das plataformas, para evitar possíveis responsabilizações pelo material compartilhado em seus ambientes digitais.

A Lei de Direitos Autorais australiana (*Copyright Act 1968*[139]) determina que um provedor de serviços que armazene conteúdo protegido por direitos autorais deve agir rapidamente para remover ou bloquear o acesso a conteúdo que infringe direitos de autor. Essa disposição é aplicável quando o provedor é informado da infração

135 A CEO da Google na Austrália, Melanie Silva, se pronunciou diversas vezes no início de 2021, através de cartas abertas e mesmo durante audiências com parlamentares australianos envolvidos na confecção da lei. Para mais informações, conferir: GOOGLE. **Mel Silva's opening statement to the Senate Economics Committee Inquiry.** Disponível em: https://blog.google/around-the-globe/google-asia/australia/mel-silvas-opening-statement. Acesso em: 19 maio 2021. GOOGLE. **Update on the News Media Bargaining Code and Google in Australia**. Disponível em: https://about.google/google-in-australia/jan-6-letter/. Acesso em: 19 maio 2021. E também: GOOGLE. **Supporting Australian journalism:** a constructive path forward An update on the News Media Bargaining Code. Disponível em: https://about.google/google-in-australia/an-open-letter/. Acesso em: 19 maio 2021.

136 BBC. Facebook blocks Australian users from viewing or sharing news. **BBC News** 18 fev. 2021. Disponível em: https://www.bbc.com/news/world-australia-56099523. Acesso em: 19 maio 2021.

137 AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Projeto de Lei para Segurança Online.** Camberra, 24 de fevereiro de 2021. Disponível em: https://www.legislation.gov.au/Details/C2021B00018. Acesso em: 17 jun. 2021.

138 AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei Para Melhorar a Segurança Online.** Camberra, 10 de setembro de 2018. Disponível em: https://www.legislation.gov.au/Details/C2018C00356. Acesso em: 17 jun. 2021.

139 AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei de Direitos Autorais.** Camberra, 01 de maio de 1969. Disponível em: https://www.legislation.gov.au/Details/C2021C00044. Acesso em: 17 jun. 2021.

ou mesmo quando fatos ou circunstâncias observadas tornem aparente que um material compartilhado representa uma infração a direitos autorais.

De maneira similar, a Lei Sobre Compartilhamento de Material Abominavelmente Violento enuncia que provedores de conteúdo devem reportar à Polícia Federal Australiana qualquer compartilhamento de "conteúdo abominavelmente violento" em seus ambientes digitais.[140] Há previsão de multas para provedores que não disponibilizarem essa comunicação em tempo "razoável" após a descoberta do conteúdo. Não há, contudo, especificação do que seria considerado um prazo razoável para essa obrigação. Há, ainda, previsão de multas ou até mesmo responsabilização penal para as plataformas que não providenciarem rapidamente a eliminação de conteúdos dessa natureza que forem facilmente acessíveis a pessoas na Austrália.[141] Nesse sentido, a lei australiana sobre conteúdo abominavelmente violento assemelha-se ao NetzDG – o que atraiu similares críticas no sentido de impor às plataformas um dever de fiscalização e julgamento que usualmente é atribuído ao Poder Público.

A Lei Para Melhorar a Segurança Online determina que as plataformas digitais devem proibir as práticas de cyberbullying dentro de seus ambientes digitais, bem como estabelecer um mecanismo para denúncia dessas práticas dentro de seus serviços.[142] Há a previsão de responsabilização das plataformas apenas caso falhem em atender a um pedido de remoção de conteúdo encaminhado diretamente pelo Comissário de e-Segurança (*e-Safety Commissioner*).[143]

O Projeto de Lei para Segurança Online entrou em tramitação durante a realização do presente estudo e reiterou diversos pontos trazidos pela Lei Para Melhorar a Segurança Online, mas expandiu enunciados e definições. O projeto de lei, por exemplo, amplia o escopo dos tipos de conteúdos que podem ser alvo de ordens de remoção pelo Comissário de e-Segurança, para incluir também material abusivo ou ofensivo direcionado a adultos, fotos íntimas cujo compartilhamento não foi consentido, entre outros.[144] Além disso, há a previsão de um prazo reduzido – de apenas 24 horas[145], comparado com o modelo atual de 48 horas[146] – para que as plataformas atendam a esses pedidos de remoção de conteúdo. Adicionalmente, o projeto de lei enuncia que o Comissário de e-Segurança pode determinar, por meio de instrumento legal, que plataformas sigam requisitos específicos para

140 Lei Sobre Compartilhamento de Material Abominavelmente Violento. Seção 474.33 (1).

141 Lei Sobre Compartilhamento de Material Abominavelmente Violento. Seção 474.34 (1) e (5).

142 Lei Para Melhorar a Segurança Online. Seção 21 (1).

143 Lei Para Melhorar a Segurança Online. Seção 44G.

144 Projeto de Lei para Segurança Online. Seção 7.

145 Projeto de Lei para Segurança Online. Seção 65 (1), (g), (i).

146 Lei Para Melhorar a Segurança Online. Seção 29 (1), (h).

garantir a segurança no meio digital[147]. Essas obrigações podem ser direcionadas a plataformas específicas ou mesmo a gêneros determinados de plataformas. Essas ordens, contudo, não representam obrigações que podem ser coercitivamente aplicadas através de processos judiciais.

A Lei de Emenda às Leis do Tesouro trouxe enunciados que visam garantir mais direitos para veículos de mídia jornalística na internet. De maneira similar a outras legislações nesse sentido aprovadas ao redor do mundo, como na Alemanha e na França, essa lei estabelece a necessidade de que plataformas digitais – como redes sociais, motores de busca, entre outros – negociem o pagamento de *royalties* com os veículos jornalísticos, para possibilitar o compartilhamento desse tipo de material protegido por direitos autorais nas plataformas online.[148] Finalmente, a lei define obrigações adicionais para as plataformas digitais, como a necessidade de informar os veículos jornalísticos sobre mudanças no funcionamento de seus algoritmos.[149] Assim, a lei impõe um dever de transparência por parte das plataformas, mesmo que não necessariamente direcionado para o usuário final, mas sim aos veículos jornalísticos que divulgam seu conteúdo nas plataformas digitais.

c) Requisitos de transparência

- **Quanto às obrigações das plataformas**
  A regulação analisada, em sua vasta maioria, não aborda pontos relativos à transparência na moderação de conteúdo online. As previsões, em regra, limitam-se às espécies de material que deve ser moderado e como deve ser realizada essa moderação.

  A exceção a essa regra, mesmo que parcial, diz respeito ao dever de transparência das plataformas de conteúdo para com os veículos jornalísticos que divulgam seu material nessas plataformas. Isso porque a Emenda às Leis do Tesouro versa sobre a necessidade de que as plataformas de conteúdo informem a mídia jornalística sobre alterações realizadas em seus algoritmos de ranqueamento de conteúdo.[150]

- **Quanto à restrição de conteúdo ilícito ou com base em políticas de comunidade**
  Não foram identificados na legislação analisada.

- **Quanto a políticas de comunidade**
  Não foram identificados na legislação analisada.

- **Quanto a relatórios de transparência**

147 Projeto de Lei para Segurança Online. Seções 45 e 46.

148 Lei de Emenda às Leis do Tesouro. Divisão 6.

149 Lei de Emenda às Leis do Tesouro. Seções 52S, 52T e 52U.

150 Lei de Emenda às Leis do Tesouro. Seções 52S, 52T e 52U.

O Projeto de Lei para Segurança Online determina que, caso haja expectativas de segurança online[151] aplicáveis a uma plataforma, o Comissário de e-Segurança pode requerer a emissão de relatórios periódicos[152]. Esses relatórios devem conter informações sobre o cumprimento das obrigações de segurança digital às quais a plataforma está submetida.

d) Obrigações quanto à moderação de conteúdo

Em termos gerais, a regulação aborda situações em que conteúdo considerado ilícito deve ser removido de plataformas online. Isso diverge quanto às previsões que dizem respeito ao grau de responsabilidade dos provedores e a sua obrigação de realizar a moderação de forma proativa ou reativa a partir de ordens emitidas por autoridades públicas.

2.3.3. Brasil

a) Contexto político e midiático

O Brasil foi um dos primeiros países do mundo[153] a elaborar uma regulação específica para o uso da internet. Em 2014, após um processo legislativo com ampla participação popular, foi promulgada a Lei 12.965/2014, conhecida como Marco Civil da Internet (MCI)[154]. Algumas situações corroboraram para a aprovação desta lei: dentre elas, destacam-se as revelações feitas pelo ex-técnico da Agência de Segurança Nacional dos Estados Unidos, Edward Snowden, sobre as práticas de espionagem do governo estadunidense sobre diversos governos latino-americanos, incluindo o Brasil e a ex-presidenta Dilma Rousseff[155]. Além disso, à época, já existia no país o Comitê Gestor da Internet no Brasil (CGI.br), criado em 1995 para "estabelecer diretrizes estratégicas relacionadas ao uso e desenvolvimento da Internet no Brasil e diretrizes para a execução do registro de Nomes de Domínio, alocação de Endereço IP (Internet Protocol) e administração pertinente ao Domínio de Primeiro Nível '.br'"[156]. Por fim,

151 Projeto de Lei para Segurança Online. Divisão 2.

152 Projeto de Lei para Segurança Online. Divisão 3.

153 CALIXTO, Dodô. Mídia internacional destaca pioneirismo do Brasil com Marco Civil da internet. **Opera Mundi.** 26 mar 2014. Disponível em: https://operamundi.uol.com.br/politica-e-economia/34544/midia-internacional-destaca-pioneirismo-do-brasil-com-marco-civil-da-internet Acesso em 19 maio 2021.

154 BRASIL. **Lei nº 12.965, de 23 de abril de 2014.** Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Diário Oficial da União. Brasília, 24 abr. 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm Acesso em 19 maio 2021.

155 SNOWDEN, Edward. **Carta aberta ao povo do Brasil**. Disponível em: https://www1.folha.uol.com.br/mundo/2013/12/1386291-leia-integra-da-carta-de-snowden-ao-brasil.shtml Acesso em: 19 maio 2021.

156 CGI - Comitê Gestor da Internet no Brasil. **Sobre.** Disponível em: https://www.cgi.br/sobre/. Acesso em 19 maio 2021.

destaca-se que o número de usuários de internet no país crescia cada vez mais, o que ampliava, também, os problemas vivenciados pelos brasileiros no ambiente virtual. O interesse empresarial e individual da população por novas tecnologias emergiu a ponto de, a partir dos anos 2010, o Brasil passar a figurar entre os principais mercados consumidores de tecnologia do mundo[157], apesar do país ainda vivenciar uma profunda exclusão digital.

Nesse contexto, o Marco Civil da Internet foi construído para oferecer uma base principiológica para interpretação do arcabouço legal já existente, bem como para julgamento dos casos concretos envolvendo internet. No âmbito desta pesquisa, moderação de conteúdo e transparência, o MCI iniciou a delimitação de algumas regras e limites tanto para a moderação quanto para a responsabilização e atuação das plataformas. Destaca-se o estabelecimento do regime de responsabilização das plataformas apenas pelo não cumprimento de ordem judicial, estabelecido no art. 19 do MCI, e que salvaguarda os intermediários quanto ao conteúdo gerado por usuário.

Em 2009, na primeira publicação do *Freedom on the Net*, o Brasil ocupava o 4º lugar no ranking dos países que garantiam mais liberdade na internet. Aos poucos, o país caiu posições sendo que ano passado chegou à posição mais baixa até então, o 21º lugar em um ranking com 65 países[158]. De acordo com o relatório, o país ocupa esta posição pela proliferação da desinformação, com a qual o atual governo brasileiro contribui, e também pelas limitações de infraestrutura para acesso à internet[159] e negligência ao letramento digital.

A expressiva quantidade de ações judiciais para remoção de conteúdos sob alegação de que ferem direitos de personalidade, especialmente partindo de políticos em período eleitoral, corrobora para os questionamentos sobre a efetividade e a legitimidade do artigo 19 do MCI. De acordo com a Associação Brasileira de Jornalismo Investigativo (Abraji), que monitora tentativas judiciais de cercear informações na web, durante as eleições de 2018, políticos brasileiros acionaram a Justiça pelo menos 503 vezes para remover conteúdos que os criticavam de alguma forma[160]. Atualmente, o regime de responsabilidade do MCI está sob análise de constitucionalidade pelo Supremo Tribunal Federal (STF), devido ao ajuizamento de

157 DAQUINO, Fernando. Brasil: um dos maiores consumidores de eletrônicos. **Tecmundo**, 15 abr. 2013. Disponível em: https://www.tecmundo.com.br/mercado/38519-brasil-um-dos-maiores-consumidores-de-eletronicos.htm Acesso em: 19 maio 2021.

158 SHALDERS, André. Como Brasil deixou de ser um dos campeões em liberdade na internet e se tornou 'parcialmente livre'. **BBC News.** 28 out. 2020. Disponível em: https://www.bbc.com/portuguese/brasil-54714957 Acesso em 19 maio 2021.

159 FREEDOM HOUSE. **Brazil**: Freedom on the Net 2020. Disponível em: https://freedomhouse.org/country/brazil/freedom-net/2020 Acesso em 7 maio 2021.

160 ABRAJI - Associação Brasileira de Jornalismo Investigativo. **Políticos tentaram ocultar informação mais de 500 vezes durante a campanha**. 21 nov. 2018. Disonível em: https://www.ctrlx.org.br/noticia/politicos-tentaram-ocultar-informacao-mais-de-500-vezes-durante-a-campanha Acesso em 19 maio 2021.

ação com pedido de indenização pela omissão do Facebook em retirar determinado perfil da plataforma após requisição de usuário[161].

b) Conteúdo da regulação

De forma similar à seção 230 da Lei de Decência das Comunicações dos Estados Unidos da América (*Communications Decency Act*), o artigo 19 do MCI estabelece o regime de responsabilidade civil que aplica sanção ao provedor apenas caso descumpra ordem judicial[162] de moderação de conteúdo (seja retirada, limitação ou outra prática de moderação). Ele conta, entretanto, com uma distinção importante: não garante isenção de responsabilidade quando um conteúdo é removido devido a políticas de comunidade (presente na lei dos EUA e chamada de "cláusula do bom samaritano")[163]. Ademais, o Marco Civil estabelece princípios já positivados pela Constituição Brasileira, como a liberdade de expressão, pluralidade e diversidade, e também dispõe sobre a proteção aos dados pessoais e garantia da neutralidade da rede, dentre outros.

Entre discursos que exigem regras mais severas de responsabilização de plataformas e usuários e discursos que exigem mais liberdade na rede para que não se instale censura privada, o Projeto de Lei (PL) 2.630/2020[164] desponta como expressão mais concreta da atenção dos diferentes setores da população brasileira sobre o tema da moderação de conteúdo e liberdade de expressão na internet. O PL, que já foi aprovado no Senado Federal e atualmente tramita na Câmara dos Deputados, estabelece regras para a moderação de conteúdo com foco nos procedimentos desta prática, além de criar um órgão responsável pela pauta, o Conselho de Transparência. O PL legitima práticas já exercidas pelas plataformas com base em suas próprias disposições, o que é denominado "autorregulação regulada" pelo projeto.

161 BRASIL. STF - Supremo Tribunal Federal. **STF discutirá validade de regra do Marco Civil da Internet sobre responsabilização de sites e redes sociais**. 2 mar. 2018. Disponível em: http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=371229&ori=1 Acesso em 19 maio 2021.

162 Esse regime é popularizado sob o termo "*judicial notice and takedown*". Neste caso, apenas após a ordem judicial é que se constitui a obrigação do intermediário em remover o conteúdo, sem que haja sua responsabilização retroativa ao início do processo que deu origem à ordem de retirada; portanto, não há um dever de indenizar, contanto que a ordem judicial seja cumprida. Ver: TOSCANO, Eliphas Neto Palitot; DE MORAIS, Annaís Moraes. O Marco Civil da Internet como novo paradigma para a responsabilidade civil dos provedores de serviços da rede: o judicial notice and takedown. In: **Direito e Novas Tecnologias II**. SIMÃO FILHO, Adalberto; PEREIRA JÚNIOR, Antonio Jorge, NASCIMENTO, Valéria Ribas do (org.). – Florianópolis : CONPEDI, 2014. Disponível em: http://www.publicadireito.com.br/artigos/?cod=93c442d40a4e0b6f Acesso em: 18 jun. 2021.

163 SOUZA, Carlos Affonso de. Brasil não precisa importar nova regra de Trump sobre redes sociais. **Tilt**, 29 maio 2020. Disponível em: https://tecfront.blogosfera.uol.com.br/2020/05/29/brasil-nao-precisa-importar-nova-regra-de-trump-sobre-redes-sociais/?cmpid=copiaecola Acesso em: 11 jun. 2021.

164 BRASIL, CÂMARA DOS DEPUTADOS. **Projeto de Lei nº 2.630, 03 jul. 2020.** (Do Sr. Alessandro Vieira). Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=node01099cclmqrw1y1s9ebm6006wp43046552.node0?codteor=1909983&filename=PL+2630/2020. Acesso em: 19 maio 2021.

O Brasil também possui o PL 21/2020[165], que trata do uso da Inteligência Artificial no Brasil – inclusive na moderação de conteúdo – e que ganhou ainda mais força após a aprovação da Estratégia Brasileira de Inteligência Artificial[166]. Em seu núcleo principal, o projeto estabelece princípios e direitos relacionados ao uso dessa tecnologia.

c) Requisitos de transparência

- **Quanto às obrigações das plataformas**
  Apesar de a regra ser de remoção apenas por ordem judicial, aplica-se exceção no caso de conteúdo sexual ou de nudez denunciado por pessoa envolvida. Nestes casos, comprovada a inércia da plataforma mesmo após notificação – judicial ou não –, ela pode ser responsabilizada subsidiariamente, segundo o Marco Civil da Internet[167]. A legislação vigente também obriga plataformas que detenham informações de contato do usuário a notificá-los dos motivos da remoção de conteúdo por ordem judicial[168]. Assim, o Marco Civil da Internet traz algumas menções à transparência e abertura nos processos envolvendo o uso da internet no Brasil. Entretanto, as obrigações de transparência ainda estão concentradas nos Projetos de Lei (PL), ou seja, ainda não integram o cenário jurídico do país. Dessa forma, a atuação dos intermediários na moderação de conteúdo encontra-se atualmente sob autorregulação.

  Como mencionado anteriormente, o principal PL que estabelece obrigações de transparência de moderação de conteúdo é o PL 2630/20. Ele propõe novas obrigações a todos os provedores com mais de 2 milhões de usuários registrados[169][170]. Uma delas é a publicização do caráter automatizado das contas que não forem geridas exclusivamente por humanos, ou seja, em regra as contas de bots devem

165 BRASIL, CÂMARA DOS DEPUTADOS.. **Projeto de Lei nº 21, 04 fev. 2020**. (Do Sr. Eduardo Bismarck). Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928&filename=PL+21/2020 Acesso em: 19 maio 2021.

166 BRASIL. Ministério da Ciência, Tecnologia e Inovações. **Estratégia Brasileira de Inteligência Artificial é publicada no DOU.** 9 abr. 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2021/04/estrategia-brasileira-de-inteligencia-artificial-e-publicada-no-dou#:~:text=A%20Estrat%C3%A9gia%20Brasileira%20de%20Intelig%C3%AAncia%20Artificial%20tem%20como%20objetivos%3A%20contribuir,e%20formar%20profissionais%20para%20o. Acesso em: 19 maio 2021.

167 Marco Civil da Internet, artigo 21.

168 Marco Civil da Internet, artigo 20.

169 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 1º.

170 Sendo que para os provedores com menor público as obrigações possuem caráter recomendatório, ou seja, não são vinculantes.

ser identificadas como tal [171][172].

Intermediários teriam, se aprovado o PL, a obrigação de remover contas automatizadas não identificadas e as contas que simulam a identidade de terceiros, que são proibidas pelo PL[173]. Esta proibição traz preocupações quanto à necessidade das plataformas checarem a verdadeira identidade de todos os seus usuários, pois o projeto não dispõe de padrões e processos para esse tipo de monitoramento. Desta forma, amplia-se a vigilância privada que os intermediários já desempenham sem estabelecer estruturas concretas para tal identificação. Apesar do Brasil possuir uma legislação específica sobre proteção de dados pessoais, a Lei 13.709/18[174], o poder de mercado de algumas plataformas demanda que a ampliação do tratamento de dados ocorra apenas em situações necessárias. Ao impor uma obrigação de identificação de usuários que amplia a coleta de dados pessoais é importante que o projeto de lei estabeleça o limite desta prática, de forma que a obrigação de identificação de usuários se restrinja à medida do necessário e não viole direitos de personalidade.

Outra obrigação estabelecida pelo PL 2630/20 é a de publicização destacada dos conteúdos pagos, seja por ser conteúdo publicitário, seja impulsionado[175]. Isso busca tornar nítida a existência de certa estrutura profissional por trás de determinados conteúdos. O projeto é ainda mais rigoroso quanto à transparência de propaganda eleitoral impulsionada, incluindo o valor gasto, identificação do anunciante, características da audiência a qual o conteúdo foi direcionado, entre outras informações. Este detalhamento é exigido pela proposta de lei considerando o amplo uso das redes sociais nas campanhas eleitorais no Brasil e o aumento da desinformação e manipulação da opinião pública durante estes períodos. Destaca-se que o Brasil já possui uma estrutura de ferramenta para campanhas eleitorais, e o PL em questão vem para estender tal transparência para o contexto digital.

Por fim, uma outra inovação institucional que se destaca no PL 2630/20 é a obrigação do Congresso Nacional de criar o Conselho de Transparência e Responsabilidade na

171 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 6º, inciso II.

172 Este foi um avanço na proposta, que inicialmente proibia completamente a existência de contas automatizadas e foi alterada para que todo esse tipo de conta seja identificado como tal, de forma a desestimular o uso de robôs para disseminação de desinformação ou manipulação da opinião pública. Isso foi feito tendo em conta a existência de diversos "bots do bem" que ajudam o cidadão a se informarem e até mesmo possibilitam o acompanhamento de políticas públicas.

173 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 6º, incisos I e II.

174 BRASIL. Lei Geral de Proteção de Dados Pessoais (LGPD). Lei 13.709, de 14 de agosto de 2018. **Diário Oficial da União**, Brasília, DF, Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 20 maio 2021

175 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 6º, inciso III.

Internet[176]. O Conselho teria como responsabilidade o acompanhamento da efetivação da lei e a complementação das lacunas de obrigações ainda existentes. Considerando o caráter multissetorial do conselho e o caráter generalista de muitas obrigações criadas pelo projeto legislativo, a criação deste órgão poderá contribuir para que as mencionadas exigências resultem no aumento da disponibilização e alcance das informações relativas à moderação de conteúdo. Se efetivamente implementado, possibilitaria que eventuais abusos sejam reparados e seus responsáveis sejam autuados.

O PL 21/2020, além de estabelecer a transparência, explicabilidade[177], responsabilização e prestação de contas[178] como princípios para o uso de inteligência artificial, também traz alguns direitos concretos para as partes interessadas no uso dessa tecnologia. Como inteligência artificial é aplicada para a moderação automatizada de conteúdo – que, como mencionado, possui diversos tópicos preocupantes relacionados a discriminação, enviesamento e erro de decisão –, os direitos incorporados pelo PL 21/2020 são essenciais. Dentre tais direitos, destaca-se o acesso a informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados pelo sistema de inteligência artificial[179]. Entretanto, a lei não estabelece parâmetros objetivos que possam garantir a efetividade deste direito, sendo essencial uma complementação legal para que as obrigações estabelecidas sejam cumpridas.

- **Quanto à restrição de conteúdo ilícito ou com base em políticas de comunidade**
  Apesar de no Brasil não existirem marcos normativos ou projetos de lei que estabeleçam situações em que determinado conteúdo deve ser restrito de forma proativa, o PL 2630/20 estabelece alguns procedimentos que devem ser seguidos caso o provedor intervenha no conteúdo gerado por terceiro. Entre essas exigências, destaca-se a obrigação de notificação ao usuário sobre tal interferência[180], incluindo informações sobre o processo envolvendo tal moderação e quais foram as medidas aplicadas. Esta exigência de notificação posterior é dispensada em alguns casos, como quando o conteúdo referir-se a crimes tipificados na Lei nº 7.716, de 5 de janeiro de 1989 ou a medida for necessária para a segurança do usuário[181]. Em casos de restrição de conteúdo ou de remoção de contas – que em alguns casos é obrigatória

176 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 25.

177 Projeto de Lei nº 21, 04 fev. 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências, artigo 6º, inciso IV.

178 Projeto de Lei nº 21, 04 fev. 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências, artigo 6º, inciso VI.

179 Projeto de Lei nº 21, 04 fev. 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências, artigo 7º,inciso II.

180 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 12, parágrafo 1º.

181 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 12, parágrafo 2º.

pelo próprio PL 2630/20 – o provedor deve oferecer ao usuário formas de recorrer à decisão[182] e direito de resposta[183].

- **Quanto a políticas de comunidade**
  Como já mencionado, o PL 2630/20 estabelece que, nos termos de uso – que também podem se referir às políticas de comunidade – os provedores devem informar como fazem a identificação dos usuários que utilizam as contas cadastradas em seus sistemas[184]. Essa identificação de contas é necessária em razão da supramencionada obrigação dos provedores de excluirem contas automatizadas não autoidentificadas como tal ou contas inautênticas. Para cumprirem com tal obrigação é necessário que as plataformas realizem a identificação de todos os usuários.

  Entretanto, somente a obrigação de informar nos termos de uso sobre os mecanismos de identificação de usuários utilizados pelas plataformas não é suficiente para transparência. Como o projeto não estabeleceu o nível de detalhamento dos procedimentos aplicados ou quaisquer outros critérios sobre quais informações acerca dessa identificação devem constar nas diretrizes do provedor, esta obrigação não implica diretamente em uma medida de transparência.

- **Quanto a relatórios de transparência**
  O PL 2630/20 propõe que as plataformas submetidas à lei divulguem semestralmente relatórios de transparência sobre os procedimentos e as decisões de tratamento de conteúdos gerados por terceiros no Brasil[185]. O PL estabelece diversas informações que devem constar nos relatórios, incluindo a quantidade de medidas de moderação de conteúdo, suas justificativas e procedimentos[186]; informações sobre o atendimento de solicitações de ordem judicial[187] e características do setor responsável pelas políticas aplicáveis a conteúdos gerados por terceiros[188].

  Além disso, o PL 21/2020 também estabelece que a União, os Estados, o Distrito Federal e os Municípios poderão solicitar aos agentes dos sistemas de inteligência

182 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 12, parágrafo 3º.

183 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 12, parágrafo 6º.

184 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 7º, parágrafo único.

185 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 13.

186 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 13, inciso II.

187 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 13, inciso IV.

188 Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet, artigo 13, inciso VII.

artificial, observadas as suas funções e justificada a necessidade, a publicação de relatórios de impacto de inteligência artificial[189]. Entretanto, diferentemente do PL 2630/2020 o projeto sobre Inteligência artificial é extremamente superficial quanto às informações que devem constar nos relatórios solicitados pelo governo.

d) Obrigações quanto à moderação de conteúdo

Atualmente o Brasil apenas estabelece a obrigação das plataformas retirarem conteúdo sexual ou nudez após denúncia do envolvido, sendo prevalecente o art. 19 do Marco Civil da Internet que responsabiliza os provedores apenas pelo não cumprimento de ordem judicial. Entretanto, o PL 2630/2020, ao propor a obrigação de identificação dos usuários que utilizam as contas registradas, também exige a realização de monitoramento ativo dos conteúdos e contas que circulam nas plataformas.

2.3.4. Índia

a) Contexto político e midiático

As notícias sobre regulação de internet na Índia apontam para uma preocupação da mídia política e dos legisladores do país em relação à falta de transparência na moderação de conteúdo. Ela é associada, majoritariamente, a desinformação, crimes contra menores de idade e aumento do discurso de ódio online com repercussões offline, como nos casos de ataques a muçulmanos[190].

No início de 2020, uma reportagem do Wall Street Journal (WSJ) divulgou conversas e relatos de funcionários do Facebook que indicaram que a empresa conscientemente se omitia em relação à restrição de conteúdo odioso publicado pelo Partido do Povo Indiano (Bharatiya Janata Party - BJP)[191] . A motivação do Facebook mencionada pelo WSJ é, principalmente, o prejuízo dos objetivos da plataforma na Índia, maior mercado da empresa[192], pela interferência nas contas de membros do parlamento. O tratamento diferenciado do Facebook para com o partido do governo foi exposto por uma funcionária da empresa à época, Sophie Zang[193]. Zang denunciou diversas

189 Projeto de Lei nº 21, 04 fev. 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências, artigo 13.

190 CHUNDURU, Aditya. Need Transparency In Facebook's Content Moderation Enforcement, Witnesses Tell Delhi's Peace And Harmony Committee. **Medianama**. 14 out. 2020. Disponível em: https://www.medianama.com/2020/10/223-need-transparency-facebook-content-moderation-delhi-peace-harmony-commiittee/. Acesso em: 10 maio 2021.

191 PURNELL, Newley; HORWITZ, JeffFacebook's Hate-Speech Rules Collide With Indian Politics. **The Wall Street Journal**, 14 ago. 2020. Disponível em: https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346. Acesso em 19 maio 2021.

192 Ibidem

193 WONG, Julia. How Facebook let fake engagement distort global politics: a whistleblower's account. **The Guardian**, 12 abr. 2021. Disponível em: https://www.theguardian.com/technology/2021/

práticas da empresa relacionadas a moderação de conteúdo seletiva e desigual.

Assim, o suposto enviesamento político e moderação de acordo com interesses econômicos das empresas escancararam a importância da transparência enquanto obrigação imposta pelo governo indiano, em propostas de co-regulação, ou a realização de um acordo corporativo entre as próprias plataformas, no modelo de autorregulação. Após o escândalo, o Comitê da Paz e Harmonia da Assembleia Legislativa de Delhi realizou várias audiências com dirigentes do Facebook para apurar o ocorrido. Na ocasião, críticas à falta de transparência na plataforma foram evidenciadas por testemunhas e sugestões foram apresentadas. Dentre elas, a obrigação de plataformas apresentarem ao governo indiano informações sobre todas as reclamações feitas a elas de forma a possibilitar o seu rastreamento e verificar a omissão ou negligência das plataformas em relação a grupos ou indivíduos específicos[194].

As potenciais arbitrariedades existentes na moderação de conteúdo que circula entre usuários indianos justificam-se por dois motivos apontados pelo WSJ. O primeiro é o desempenho das funções de execução de políticas das plataformas e a função de lobby pela mesma equipe das empresas. Esta situação ocorre no time indiano de políticas públicas do Facebook, ou seja, a equipe da plataforma responsável pelas regras de moderação de conteúdo e a equipe responsável por desenvolver boas relações com o governo é a mesma[195]. Dessa forma, conflitos de interesses são inevitáveis e dificultam a aplicação imparcial das regras previamente definidas. O segundo motivo apontado para questionamentos quanto à moderação de conteúdo na Índia é o controle governamental antidemocrático e sucessivas ameaças feitas pelo governo indiano – especificamente pela ala conservadora do governo – aos funcionários da plataforma. Recentemente, após a atuação do Twitter na moderação de conteúdo referente a protestos de fazendeiros indianos, funcionários do Twitter e Facebook foram ameaçados de prisão caso não removessem as contas e publicações solicitadas pelo governo[196].

Outro fator de movimentação das discussões na Índia acerca da transparência na moderação de conteúdo é o embate geopolítico existente no continente. As disputas

apr/12/facebook-fake-engagement-whistleblower-sophie-zhang. Acesso em 16 jun. 2021.

194 ROSHAN, Utkarsh. After India and Australia Controversies, Facebook Changes Terms and Conditions Again. **The Citizen,** 16 set. 2020. Disponível em: https://www.thecitizen.in/index.php/en/NewsDetail/index/9/19369/After-India-and-Australia-Controversies-Facebook-Changes-Terms-and-Conditions-Again. Acesso em 19 maio 2021.

195 KUMAR, Ankit; PATEL, Anand. Enforce policies regardless of political position: Under heat, Facebook denies ties with BJP in India. **India Today**. 16 ago. 2020. Disponível em: https://www.indiatoday.in/india/story/enforce-policies-regardless-political-position-under-heat-facebook-bjp-india-1711873-2020-08-16. Acesso em 19 maio 2021.

196 HORWITZ, Jeff; PURNELL, Newley. ndia Threatens Jail for Facebook, WhatsApp and Twitter Employees. **The Wall Street Journal**, 5 mar. 2021. Disponível em: https://www.wsj.com/articles/india-threatens-jail-for-facebook-whatsapp-and-twitter-employees-11614964542. Acesso em 19 maio 2021.

entre China e Estados Unidos da América que envolvem acusações de espionagem e uso ilegal de dados pessoais também repercutiram nos vizinhos asiáticos, incluindo a Índia[197]. Em 2020, a Índia baniu permanentemente 58 aplicativos chineses, incluindo o TikTok, por suposta ameaça à soberania nacional. Já o Paquistão baniu o aplicativo por suposta falha da empresa em moderar conteúdos indecentes e imorais[198]. Essas ações estimularam a ByteDance, companhia proprietária do TikTok, a criar um centro de aconselhamento para moderação de conteúdo na Ásia e Pacífico, o que se aproxima de um modelo de autorregulação, porém, com uma composição multissetorial e atuação de representantes do governo[199].

Na mesma linha, a criação do Oversight Board do Facebook[200], que também possui um membro da Índia[201], mostra uma tendência das empresas de reformularem a autorregulação. Agora, ela passa a incluir representantes de outras entidades e setores na tomada de decisão sobre moderação de conteúdo.

As interferências governamentais sobre a atuação da plataforma na Índia chamam atenção para um segundo ponto relacionado com os riscos ao exercício de direitos digitais: a dificuldade em utilizar a transparência para gerar responsabilização. Já se mencionou como transparência e responsabilização se relacionam e, principalmente, como a transparência é um meio para efetivar direitos e não um fim em si mesmo. Portanto, as sucessivas interferências indevidas do governo na atuação das plataformas demonstram que, ainda que a Índia estabelecesse obrigações de transparência na moderação de conteúdo, não é garantido que essas obrigações resultariam em um ambiente informacional mais livre e saudável.

b) Conteúdo da regulação

A legislação indiana foca em regular intermediários, especialmente os de maior público, alterando o regime de responsabilidade sobre conteúdo de usuário em algumas situações.

197 GOODFELLOW, Jessica. Pakistan reverses TikTok ban. **Campaign Asia,** 20 out. 2020. Disponível em: https://www.campaignasia.com/article/pakistan-reverses-tiktok-ban/464364. Acesso em 19 maio 2021.

198 KLEINA, Nilton. Índia bane TikTok e mais 58 apps chineses por 'ameaça à soberania'. **Tecmundo**, 29 jun. 2020. Disponível em: https://www.tecmundo.com.br/software/154648-india-bane-tiktok-58-apps-chineses-ameaca-soberania.htm Acesso em 19 maio 2021.

199 AGARWAL, Surabhi. Officials debate whether India should have its own social media content moderation rules. **The Economic Times**. 1 set. 2020. Disponível em: https://news.google.com/articles/CAIiELSW6qmuRu6lx_AX0NtvBJkqGQgEKhAIACoHCAow2pqGCzD954MDMMLNigY?hl=en-IN&gl=IN&ceid=IN%3Aen. Acesso em 19 maio 2021.

200 HARRIS, Brent. Establishing Structure and Governance for an Independent Oversight Board. **Facebook.** 17 set. 2019. Disponível em: https://about.fb.com/news/2019/09/oversight-board-structure/. Acesso em 19 maio 2021.

201 Trata-se do ativista Sudhir Krishnaswamy. Ver: TECH DESK. Who is Sudhir Krishnaswamy from Facebook's oversight board? **The Indian Express**. 8 maio 2020. Disponível em: https://indianexpress.com/article/technology/tech-news-technology/sudhir-krishnaswamy-facebook-oversight-board-6398217/ Acesso em: 18 jun. 2021.

As Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021 (cujo nome em inglês é *Information Technology – Guidelines For Intermediaries And Digital Media Ethics Code – Rules, 2021*)[202] suprimem o *Information Technology Rules* de 2011, que estabelecia algumas obrigações aos intermediários para moderação de conteúdo, agora incorporadas em parte pela nova lei.

Apesar de direcionada a intermediários significativos – aqueles com mais de 5 milhões de usuários registrados –, a legislação possibilita que o Ministério de Eletrônica e Tecnologia da Informação requeira que qualquer intermediário que disponibilize ferramentas para publicação de conteúdo que coloque em risco a soberania da Índia, a relação com outros países ou a ordem pública também seja submetido às obrigações desta lei[203]. A discricionariedade que a lei oferece ao governo para definir quais entidades estão submetidas às obrigações é um dos primeiros pontos de atenção para que potenciais arbitrariedades e conformidades discriminatórias sejam estabelecidas[204].

A Índia possui um regime de responsabilidade limitada das plataformas, segundo o qual os intermediários não respondem judicialmente por conteúdo publicado por terceiro. Ainda assim, as novas sanções das Regras de Tecnologia da Informação enrijecem as responsabilidades dos intermediários ao estabelecer sanções, inclusive penais, em caso de violação. Desta forma, apesar do regime de responsabilidade em relação ao conteúdo publicado por terceiro se manter, ao estabelecer novas obrigações de monitoramento de conteúdo, a legislação indiana amplia o escopo de ações que as plataformas podem tomar.

Quanto às sanções estabelecidas pelo novo marco normativo, destaca-se que a legislação é vaga e menciona apenas que os intermediários serão responsabilizados. Não são determinados quaisquer parâmetros para a aplicação das sanções e tampouco é mencionado quais são os tipos de sanções cabíveis. A inexistência de critérios e de tipos sancionatórios amplia a zona de discricionariedade dos responsáveis pela aplicação das sanções – o que também estende as chances de imposição de decisões arbitrárias e excessivas.

Destaca-se que a lei não aborda apenas a atuação dos intermediários, mas também possui um capítulo específico relacionado às plataformas de distribuição de conteúdos – que vão desde plataformas de *streaming* como Amazon Prime e Netflix até veículos jornalístico como *The New York Times* – aos quais uma série de limitações

202 INDIA. MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY. **Regras de Tecnologia da Informação** – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021. Nova Delhi, 25 de Fevereiro de 2021. Disponível em: http://egazette.nic.in/WriteReadData/2021/225464.pdf. Acesso em: 19 maio 2021.

203 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 6, subseção 1.

204 IFF - Internet Freedom Foundation. **Deep dive**: How the intermediaries rules are anti-democratic and unconstitutional. 21 fev. 2021. Disponível em: https://internetfreedom.in/intermediaries-rules-2021/. Acesso em: 19 maio 2021.

à publicação de certos conteúdos é estabelecida[205]. Nesse sentido, uma estrutura robusta composta por um órgão do governo responsável por decidir disputas entre usuários e editores – e com poderes para decidir pela retirada ou modificação de conteúdo – é estabelecida pela lei com objetivo de oferecer controle ao conteúdo circulado online.

c) Requisitos de transparência

- **Quanto às obrigações das plataformas**
  As plataformas são obrigadas, pelas Regras de Tecnologia da Informação, a armazenar informações por períodos variados, como conteúdos denunciados ou notificados pelo governo como ilícitos[206] e informações de registro de usuário[207]. Além de não haver uma obrigação de transparência acerca desse armazenamento de informações, esta obrigação deixa vulneráveis os dados pessoais dos usuários indianos, considerando que o país não possui uma legislação de proteção de dados específica ou qualquer outra supervisão contra o vigilantismo[208].

  Além disso, os intermediários são obrigados a disponibilizar mecanismo de denúncia e reclamação por violações às regras da plataforma[209], sendo que as medidas tomadas pelo intermediário após a contestação devem ser informadas ao reclamante[210]. Destaca-se a obrigação de identificação de conteúdos que resultam em lucro para alguma parte (autor ou intermediário) como conteúdo patrocinado, comercializado ou outra classificação[211].

  Os intermediários também devem fornecer meios e recursos para que os usuários verifiquem suas contas e tenham suas contas publicamente identificadas como verificadas[212]. Apesar de esta ser uma medida justificada pelo objetivo de coibir a existência de robôs e contas falsas ou maximizar a identificação de usuários

205 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, parte III.

206 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, parte II, seção 3, subseção 1, cláusula g.

207 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 3, subseção 1, cláusula h.

208 IFF - Internet Freedom Foundation. **Deep dive**: How the intermediaries rules are anti-democratic and unconstitutional. 21 fev. 2021. Disponível em: https://internetfreedom.in/intermediaries-rules-2021/. Acesso em: 19 maio 2021.

209 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 3, subseção 2 e seção 4, subseção 6.

210 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 6.

211 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 3.

212 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 7.

maliciosos, é possível que o caráter voluntário dessa identificação se torne obrigatório. Neste cenário, usuários que dependem do anonimato e de maior privacidade – como dissidentes políticos e ativistas – podem ser prejudicados[213].

As Regras de Tecnologia da Informação estabelecem requisitos específicos para manifestação de conteúdos de cunho jornalístico ou de atualidades, incluindo a obrigação de que editores ofereçam detalhes adicionais aos intermediários sobre sua atuação online para que possam ser acompanhados pelo Ministério da Informação e Radiodifusão (*Ministry of Information and Broadcasting*)[214]. Os editores identificados como tal podem receber uma marca ou sinal correspondente a sua identificação e que seja publicamente visível.

No âmbito das obrigações sobre moderação de conteúdo, a lei estabelece que os intermediários utilizem técnicas para detecção de conteúdo de abuso infantil inclusive por meio de análise automatizada. Apesar de estabelecer aos intermediários a obrigação de revisão humana das decisões automatizadas[215], o enviesamento, descriminação e potenciais erros causados pelo uso de automatização na moderação de conteúdo demanda maior transparência e instrumentos de acompanhamento das inferências realizadas por robôs. A legislação não estabelece quaisquer mecanismos de transparência algorítmica, auditoria de sistemas ou outros instrumentos sendo mais um ponto de risco a diversos direitos humanos.

- **Quanto a restrição de conteúdo ilícito ou com base em políticas de comunidade**
  O arcabouço jurídico indiano permite que os intermediários restrinjam conteúdo ou removam os acessos do usuário à plataforma tanto por violação das regras da empresa quanto por violação à legislação indiana. Nada obstante, o intermediário deverá informar periodicamente seus usuários sobre potenciais sanções [216].

  Além disso, a remoção de conteúdo ou links só pode ser realizada desde que, antes da remoção, o usuário diretamente afetado pela moderação seja informado das razões que a motivaram[217]. A lei define "usuário diretamente afetado" como aquele que "criou, carregou, compartilhou, disseminou ou modificou informações, dados ou link de comunicação". Ademais, o intermediário deve oferecer ao usuário oportunidade razoável de defender seu conteúdo e solicitar a reversão da decisão de remoção[218]. Os mecanismos de resolução de disputas oferecidos aos usuários devem

213 Ibidem

214 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 5.

215 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 4.

216 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, parte II, seção 3, subseção 1, cláusula c.

217 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 8, cláusula a.

218 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia

ser constantemente supervisionados pelo *Resident Grievance Officer*[219].

- **Quanto a políticas de comunidade**
  Não há obrigações.

- **Quanto a relatórios de transparência**
  É obrigatório, pelas Regras de Tecnologia da Informação, fazer publicações mensais de relatórios de conformidade com a legislação que informem detalhes das reclamações recebidas, as medidas tomadas a respeito e o número de conteúdo retirados ou restringidos após monitoramento proativo de conteúdo. A legislação não é clara sobre quais "reclamações" devem constar nos relatórios de transparência. Assim, não é possível assegurar que requisições governamentais sejam publicizadas, assim como o são as solicitações de usuários comuns.

d) Obrigações quanto a moderação de conteúdo

As Regras de Tecnologia da Informação estabelecem normas gerais de uso das plataformas que devem ser incorporadas por todos os intermediários, estabelecendo os limites materiais e concretos à circulação de conteúdo online[220]. É obrigatório publicizar os termos de uso e/ou as políticas de comunidade, tanto em sites quanto aplicativos das mídias sociais. Eles devem conter, dentre suas informações, as regras de uso a serem aplicadas por todos os intermediários. Destacam-se a proibição de conteúdo difamatório, pornográfico, danoso a crianças, desinformativo ou que ameace a segurança nacional.

Especialmente quanto a conteúdos sexuais ou de nudez, a lei obriga os intermediários a retirarem o conteúdo mediante denúncia dentro de 24h[221]. A lei estabelece obrigação de monitoramento ativo – especialmente acerca de conteúdo relacionado à violência sexual – mas estabelece que esse monitoramento deve ser realizado de forma proporcional à liberdade de expressão e outros direitos humanos[222].

Digital 2021, seção 4, subseção 8, cláusula b.

219 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 8, cláusula c.

220 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, parte II, seção 3, subseção 1, cláusula b.

221 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 3, subseção 2, cláusula b.

222 Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021, seção 4, subseção 4.

2.3.5. Quênia

a) Contexto político e midiático

A democratização do Quênia coincide com o desenvolvimento de tecnologias de informação e comunicação digitais, tendo o país passado ao multipartidarismo nos anos 1990[223]. Desde essa transição na organização política, o potencial dessas formas de comunicação foi visto pelos agentes políticos (candidatos e membros de partidos) como instrumento. Isso pode justificar o fato de muitas das iniciativas sobre regulação de conteúdo online envolverem o contexto eleitoral.

O cenário regulatório do Quênia é voltado ao controle de conteúdo potencialmente ofensivo a grupos ou a indivíduos com base no pertencimento a grupos. Isso é permeado por questões interétnicas que há muito desencadeiam conflitos políticos no país e sobre as quais se tem grande preocupação desde as eleições de 2007 – marcadas por uma grande polarização entre os dois principais candidatos (Kibaki e Odinga)[224] [225] e instrumentalização da mídia. Além disso, existem práticas, adotadas abertamente, de controle prévio/censura de conteúdo e a criminalização de conteúdos considerados indesejados, como exemplificado mais adiante.

Potencialmente como reflexo de seu contexto sociopolítico, marcado por disputas interétnicas, o país aplica transparência sobre restrições a conteúdo, mas as práticas em si não se orientam por proporcionalidade e critérios de livre expressão. Isso confere ao Quênia uma avaliação negativa em relação à liberdade comunicacional pelo relatório *Freedom on The Net*, por exemplo[226]. Alguns eventos, narrados a seguir, se destacam na configuração desse cenário.

Em 2007, durante as eleições, as mídias tradicionais foram um dos âmbitos nos quais afloraram conflitos interétnicos comuns ao cenário político da época.Desde então, esse tipo de conteúdo vem sendo objeto de preocupação. A reforma constitucional de

223 NDAVULA, John; MUENI, Joy. New Media and Political Marketing in Kenya: The Case of 2013 General Elections. **International Journal of Arts and Commerce**. 3. 2014. p. 72. Disponível em: https://www.researchgate.net/publication/316189367_New_Media_and_Political_Marketing_in_Kenya_The_Case_of_2013_General_Elections Acesso em: 18 jun. 2021

224 "A vitória de Kibaki foi contestada pelo partido opositor, e gerou uma série de revoltas e episódios de violência, opondo diferentes etnias. Vale destacar que Kibaki e Odinga pertencem a grupos étnicos historicamente rivais – os Kikuyu e os Luo, respectivamente – e isso teve um peso nas eleições de 2007[...]". Em: ONIAS, Clarice Barbosa; RIBEIRO, Pedro Feliú. O impacto da presença chinesa na elite política do Quênia. **12º Encontro da ABCP**. 2020. Disponível em: https://cienciapolitica.org.br/system/files/documentos/eventos/2021/01/impacto-presenca-chinesa-elite-politica-quenia-3022.pdf Acesso em: 14 jun. 2021. p. 9

225 COOKE, Jennifer. Background on the Post-Election Crisis in Kenya. **Blog CSIS - Center for Strategic & International Studies**. 6 ago. 2009. Disponível em: https://www.csis.org/blogs/smart-global-health/background-post-election-crisis-kenya Acesso em: 14 jul. 2021.

226 FREEDOM HOUSE. Kenya: **Freedom on the Net 2020**. Disponível em: https://freedomhouse.org/country/kenya/freedom-net/2020 Acesso em 7 maio 2021.

2010, que consolidou algumas mudanças ocorridas, como o pluripartidarismo, e tem em seu rol o direito à liberdade de expressão[227], foi impulsionada por esse contexto.

Também existem, desde 2017, diretrizes oficiais sobre propagação de conteúdo político. As eleições envolveram violência e posterior suspensão de serviços de mídia devido à potencial transmissão de uma posse presidencial promovida pela oposição, que declarava inválidos os resultados da recontagem dos votos[228]. As regras foram elaboradas pela Comissão Nacional de Coesão e Integração e pela Autoridade de Comunicações (*National Cohesion and Integration Commission - NCIC*)[229].

As Diretrizes da NCIC (*NCIC Guidelines*, no original) proíbem mensagens políticas que contenham linguagem profana, obscena, confusa, enganosa, intimidatória, abusiva ou ofensiva; também requerem que administradores de páginas de mídias sociais moderem e controlem o conteúdo e as discussões geradas na plataforma e dá poderes, a provedores de serviços de telefonia móvel, de bloqueio de transmissão de mensagens políticas em massa que não se adequem às diretrizes; esse tipo de mensagem requer prévia aprovação do serviço ou da NCIC.

Nesse sentido, houve pedidos de autoridades para que usuários comuns e jornalistas removessem conteúdo de seus perfis e sites, com casos públicos quanto a postagens sobre personalidades políticas em 2017 e 2018[230] [231]. Também em 2018, alguns vídeos (de conteúdo LGBTQIA+, sendo que homossexualidade é criminalizada no Quênia[232] [233]) foram banidos pela NCIC – notadamente, o filme Rafiki[234], que teve

227 RUTENBERG, Isaac; SUGOW, Abdulmalik. Regulation of the Social Media in Electoral Democracies: A Case of Kenya. **SOAS Law Journal**, v. 7, p. 301, 2020. Disponível em: https://www.cipit.strathmore.edu/wp-content/uploads/2020/06/Journal-7.pdf Acesso em: 18 jun. 2021. p. 321.

228 NYABOLA, Nanjala. Putting Kenya's media shutdown in context. **Aljazeera**. 2 fev. 2018. Disponível em: https://www.aljazeera.com/opinions/2018/2/2/putting-kenyas-media-shutdown-in-context/ Acesso em 10 maio 2021.

229 NCIC - National Cohesion and Integration Commission. **Guidelines on prevention of dissemination of undesirable bulk and premium rate political messages and political social media content via electronic communications networks** (NCIC Guidelines). jul. 2017. Disponível em: https://ca.go.ke/wp-content/uploads/2018/02/Guidelines-on-Prevention-of-Dissemination-of-Undesirable-Bulk-and-Premium-Rate-Political-Messages-and-Political-Social-Media-Content-Via-Electronic-Networks-1.pdf Acesso em 19 maio 2021.

230 MURAYA, David. Blogger Robert Alai Arrested For Posting Gory Photos. **Capital News**. 18 jun. 2019. Disponível em: https://www.capitalfm.co.ke/news/2019/06/blogger-robert-alai-arrested-for-posting-gory-photos/ Acesso em: 7 maio 2021.

231 OMONDI, Ian. Another suspect arrested for allegedly sending Robert Alai photos of dead officers. **Citizen Digital**. 19 jun. 2019. Disponível em: https://citizentv.co.ke/news/another-suspect-arrested-for-allegedly-sending-robert-alai-photos-of-dead-officers-259094/. Acesso em: 7 maio 2021.

232 A relação entre pessoas do mesmo sexo é criminalizada no país, com pena de prisão. LOC - The Law Library of Congress. **Laws on Homosexuality in African Nations**. fev. 2014. Disponível em: https://www.loc.gov/law/help/criminal-laws-on-homosexuality/homosexuality-laws-in-african-nations.pdf Acesso em 14 jun. 2021.

233 ABRIL. Tribunal do Quênia mantém criminalização da homossexualidade. **Veja,** 24 maio 2019. Disponível em: https://veja.abril.com.br/mundo/tribunal-do-quenia-mantem-criminalizacao-do-homossexualismo/ Acesso em: 14 jun. 2021.

234 A restrição imposta impede "distribuição, exibição ou transmissão do filme em qualquer lugar

proibidas a distribuição, a transmissão ou a posse de mídia contendo o filme. Seu banimento foi suspenso apenas por alguns dias para ser considerada sua indicação ao Oscar[235 236]. Em 2019 e 2020, houve episódios de remoção de vídeos do YouTube de um grupo queniano a pedido de autoridades governamentais, sob alegação de promover atividades criminosas e abusivas contra mulheres e crianças[237]. Segundo notícias, grande parte do público reconheceu que as músicas e vídeos promoviam glorificação do estupro de maneira explícita[238].

À luz desses fatos, em 2020, a Suprema Corte decidiu que a Lei de Cibercrimes e Uso Indevido de Computadores (*Computer Misuse and Cibercrimes Act*, no original)[239], que fixa pena de prisão de até 10 anos por disseminação de informação falsa (entre outros conteúdos considerados nocivos ou ofensivos), não fere a liberdade de expressão. Um trecho da decisão[240] permite constatar que foi considerada legítima a restrição a atos de circulação de informação falsa por ser o meio considerado adequado para coibir este tipo de prática, sem que houvesse outras formas. O raciocínio decisório fundamenta-se no interesse em proteger a ordem pública. Ainda neste mesmo ano, blogueiros foram presos segundo a Seção 23 desta lei, por fazerem postagens

da República do Quênia", e alerta que "qualquer pessoa em posse do filme comete ilícito". KFCB - Kenya Film Classification Board. Statement by the Kenya Film Classification Board Chief Executive Officer Dr. Ezekiel Mutua, MBS on the banning of the film "Rafiki" issued on 27th april 2018. **KFCB**. 27 abr. 2018. Disponível em: http://kfcb.co.ke/wp-content/uploads/2018/04/CEO-STATEMENT-ON-RAFIKI-RESTRICTION-27-4-2018.pdf Acesso em: 7 maio 2021.

235 KIMUYU, Hillary; KIPLAGAT, Sam. Court lifts ban on Rafili, film to run for an Oscar. **Nation**. 28 jun. 2020. Disponível em: https://nation.africa/kenya/news/court-lifts-ban-on-rafiki-film-to-run-for-an-oscar-89512 Acesso em: 07 maio 2021.

236 Ver também, sobre a suspensão do banimento ter valido apenas por alguns dias na capital Nairóbi, a fim de permitir a elegibilidade ao Oscar: COOPER, Mariah. ARTS & ENTERTAINMENT Kenya lifts ban on lesbian film 'Rafiki,' giving it Oscar eligibility. **Washington Blade**. 21 set. 2018. Disponível em: https://www.washingtonblade.com/2018/09/21/kenya-lifts-ban-on-lesbian-film-rafiki-for-oscar-consideration/. Acesso em: 07 maio 2021.

237 MUTUA, Ezekiel. Google Takes Down Ethic's Raunchy Tarimbo Song Despite Apology. **Daily Active**. 9 set. 2019. Disponível em: https://www.dailyactive.info/2019/11/09/google-takes-down-ethics-raunchy-tarimbo-song-despite-apology/. Acesso em 19 maio 2021.

238 OBIERA, Gladys Mokeira. Ethic Entertainment's Tarimbo ban - Why and fan reactions. [out. 2020]. **Tuko**. Disponível em: https://www.tuko.co.ke/322781-ethic-entertainments-tarimbo-ban-why-fan-reactions.html. Acesso em: 07 maio 2021.

239 QUÊNIA. Computer Misuse and Cybercrimes Act. **Kenya Gazette Suplement**. Nairobi, 16 maio 2018. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/ComputerMisuseandCybercrimesActNo5of2018.pdf . Acesso em 19 maio 2021.

240 "Algumas das informações falsas resultaram em medo e pânico entre os membros do público e podem criar caos, incerteza e uma ameaça à segurança nacional do país; situação que pode não só aumentar, especialmente durante o período eleitoral, mas também se tornar emotiva. Acho, portanto, que é necessário estabelecer uma lei para regular e controlar a difusão de informações falsas que podem representar uma ameaça à segurança nacional do país, especialmente em um país tão fragmentado como o nosso." Traduzido do original: KENYALAW. Bloggers Association of Kenya (BAKE) v Attorney General & 3 others; Article 19 East Africa & another (Interested Parties) [2020] eKLR. Petition n. 206 of 2019. **Judgment**. Disponível em: http://kenyalaw.org/caselaw/cases/view/191276/ Acesso em: 15 jun. 2021. par. 45

alertando sobre a COVID-19 que foram consideradas desinformativas[241 242].

Ainda, registra-se que Main Quiai, queniana com experiência em *advocacy* por direitos humanos e ex-Relatora Especial da ONU foi escolhida para integrar a instância recursal criada pelo Facebook acerca de moderação de conteúdo[243]. Isso também significa que existem importantes atores da área no país, e que se podem esperar mais discussões no Quênia acerca de moderação de conteúdo, transparência e livre expressão.

b) Conteúdo da regulação

As respostas penais e criminalizantes são amplamente adotadas como solução a questões sobre conteúdo potencialmente conflituoso. Não existe uma estratégia com foco nos sistemas de disseminação de conteúdo, mas sim voltada ao conteúdo em si, o que é criticado por alguns estudos[244 245].

As principais leis que tratam de moderação de conteúdo no Quênia são a Lei de Cibercrimes e Uso Indevido de Computadores de 2018, que criminaliza propagação de conteúdo falso, pornografia infantil e pirataria; a Lei de Filmes e Peças de Teatro (*Films and Stage Plays Act*) de 2012[246], que regula a necessidade de autorização do Conselho de Classificação de Filmes do Quênia (*Kenyan Film Classification Board* - KFCB) para publicar certos tipos de conteúdo em alguns horários (entre eles, conteúdo contendo relacionamentos LGBTQIA+, que é proibido no país); e também as Diretrizes da NCIC sobre propagação de conteúdo político, mais direcionadas ao tópico político e sob fiscalização da Comissão Nacional de Integração e Coesão.

241 Notadamente, o blogueiro Cyprian Nyakundi, que foi processado criminalmente por postar sobre as causas da morte de um oficial da agência governamental KRA - *Kenya Revenue Authority*, e o blogueiro Robert Alai, que postou sobre a situação de propagação do coronavírus no Quênia, alegando que o governo escondia informações quando reportava ter apenas 15 casos em março de 2020. Ver: MAOMBO, Sharon. DCI declares blogger Cyprian Nyakundi a wanted man **Star**. 21 abr. 2020. Disponível em: https://www.the-star.co.ke/news/2020-04-21-dci-declares-blogger-cyprian-nyakundi-a-wanted-man/. Acesso em: 7 maio 2021.

242 KIMUYU, Hillary. Blogger Alai charged for publishing 'alarming' claims on coronavirus. Nairobi News. 23 mar 2020. Disponível em: https://nairobinews.nation.co.ke/news/blogger-alai-charged-for-publishing-alarming-claims-on-coronavirus. Acesso em: 7 maio 2021.

243 INGRAM, David. Facebook names 20 people to its 'Supreme Court' for content moderation. **NBC News**. 6 maio 2020. Disponível em: https://www.nbcnews.com/tech/tech-news/facebook-names-20-people-its-supreme-court-content-moderation-n1201181. Acesso em: 07 maio 2021.

244 RUTENBERG, Isaac; SUGOW, Abdulmalik. Regulation of the Social Media in Electoral Democracies: A Case of Kenya. **SOAS Law Journal**, v. 7, p. 301, 2020. Disponível em: https://www.cipit.strathmore.edu/wp-content/uploads/2020/06/Journal-7.pdf Acesso em: 18 jun. 2021.

245 BOWERS, John; ZITTRAIN, Jonathan L. Answering impossible questions: Content governance in an age of disinformation. **The Harvard Kennedy School (HKS) Misinformation Review**. 10 fev. 2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3520683 Acesso em 15 jun. 2021.

246 Quênia. Films and stage plays act. **National Council for Law Reporting**. 2012. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/FilmsandStagePlaysAct_Cap222.pdf. Acesso em 17 jun. 2021.

Além disso, com a preocupação crescente acerca da polarização potencializada pelas mídias sociais, tramita no parlamento o Projeto de Emenda da Lei de Informações e Comunicações (*Information and Communications Act - Amendment Bill*, no original)[247] de 2019, que propõe algumas medidas específicas. Entre elas, estão: 1. a necessidade de licença para estabelecer plataforma de mídia social, com escritório físico no país; 2. dever do usuário se certificar de que seu conteúdo seja justo, preciso, não-enviesado e não ofenda quem o receberá; 3. dever do administrador em grupos de redes sociais notificar o licenciado de suas intenções de formar um grupo na plataforma, bem como controlar conteúdo e discussões indesejáveis; 4. a criação de um registro de blogueiro[248] por uma comissão designada, sendo ilícito possuir um blog sem uma licença.

c) Requisitos de transparência

- **Quanto às obrigações das plataformas**
  Há obrigação de monitoramento ou filtro proativo de conteúdo, especialmente em relação a mensagens políticas e sua disseminação em massa por provedores via internet móvel[249], imposta pelas Diretrizes da NCIC. Essa regulação impõe que um serviço provedor de conteúdo interessado em disseminar informação política deve requerer uma autorização para o provedor de serviço de telefonia móvel. Os servidores também devem enviar o conteúdo para análise da NCIC caso não se considerem capazes de determinar a adequação da disseminação daquela mensagem, sendo que a NCIC deve responder em até 24h. Se uma plataforma vetar a disseminação de uma mensagem política, o que pode ser feito com base nas Diretrizes, deve notificar o requerente em até 18h depois do pedido, além de apresentar os motivos.

  A preocupação é limitada a apenas um aspecto unidirecional e privado de transparência com o próprio usuário que sofre os efeitos da moderação. Isto é, somente esta pessoa será comunicada no caso de negativa sobre envio de mensagem política em massa. Entretanto, as obrigações são limitadas aos sujeitos envolvidos, isto é, não é necessário que a decisão de vetar a disseminação seja disponibilizada ao público. Isso prejudica o amplo debate sobre os critérios utilizados e também

247 QUÊNIA. Information and Communications Act (Amendment Bill). 2019. **Kenya Gazette Suplement.** 24 jul. 2019, Nairobi. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/bills/2019/TheKenyaInformationandCommunication__Amendment_Bill_2019_NA_Bills_No._61.pdf Acesso em 14 jun. 2021.

248 Blogueiro, segundo o projeto de lei, é quem faz"blogging", que consiste em coletar, escrever, editar ou apresentar notícias ou artigos de notícias em plataformas de redes sociais ou na internet ("means collecting, writing, editing and presenting of news or news articles in social media platforms or in the internet"). Ver: QUÊNIA. Information and Communications Act (Amendment Bill). 2019. **Kenya Gazette Suplement.** 24 jul. 2019, Nairobi. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/bills/2019/TheKenyaInformationandCommunication__Amendment_Bill_2019_NA_Bills_No._61.pdf Acesso em 14 jun. 2021. Parte I, seção 2.

249 Diretrizes da NCIC, Parte I, Cláusula 7.2

dificulta a fiscalização do cumprimento de um fornecimento suficiente de justificativa para barrar o fluxo de informação. Terceiros independentes terão dificuldade para analisar os casos de restrição, suas razões e circunstâncias, o que abre margem para a censura de mensagens.

- **Quanto à restrição de conteúdo ilícito ou com base em políticas de comunidade**
  Não há requisitos específicos para restrição de conteúdo em geral, apenas em situação que envolve conteúdo político enviado em massa. Existe obrigação de obter uma autorização do provedor de internet móvel[250].

  Algumas categorias de conteúdo ilícito estão na Lei de Cibercrimes e Uso Indevido de Computadores, de 2018. Elas abarcam definições genéricas que criminalizam discurso de ódio, ofensa à reputação e disseminação de informações falsas[251]. Esses dispositivos sofrem questionamentos e críticas de entidades da sociedade civil[252][253], por sua ampla abrangência, que permite a criminalização com alto nível de discricionariedade.

  Ainda que não haja responsabilidade legal das plataformas pela remoção ou controle desse tipo de conteúdo, há uma lacuna de obrigações de transparência. A ausência de delimitação nítida sobre quais conteúdos uma categoria abarca, que também vem sem qualquer obrigação de discriminar os critérios para a análise de ilicitude de um conteúdo, impede que se possa avaliar medidas eventualmente adotadas.

- **Quanto a políticas de comunidade**
  Não foram identificados na regulação analisada.

- **Quanto a relatórios de transparência**
  Não foram identificados na regulação analisada.

d) Obrigações quanto à moderação de conteúdo

A regulação queniana volta-se a diversas medidas de controle de alguns tipos de conteúdo, impondo limitações abstratas ao discurso online, como as da Lei de cibercrimes ou das Diretrizes da NCIC. Entretanto, via de regra, elas não incidem sobre a moderação de conteúdo feita pelas plataformas. O nível no qual operam é o

250 Diretrizes da NCIC, Parte I, Cláusula 7.

251 Lei de Cibercrimes e Uso Indevido de Computadores, Seção 22.

252 MUYA, Catherine. Policy Analisys on the Kenya Information and Communications Act. **Lawyers Hub.** 28 out. 2019. p. 18. Disponível em: https://lawyershub.org/media/Policy_Analysis_on_Kenya_Information_and_Communcation_Amendment_Act_1_UXKIfMV.pdf. Acesso em: 10 maio 2021.

253 CIPESA - Collaboration on International ICT Policy for East and Southern Africa. State of Internet **Freedom in Kenya 2019:** mapping Trends in Government Internet Controls, 1999-2019**.** jan 2020. p. 20. Disponível em: https://www.opennetafrica.org/?wpfb_dl=100. Acesso em: 10 maio 2021.

da autoridade judicial, que pode restringir conteúdo e mesmo a liberdade de quem o dissemina.

No campo das regulações em trâmite, o Projeto de Emenda da Lei de Informações e Comunicações, ao propor que alguns agentes sejam responsabilizados por conteúdo de terceiros, pode ter reflexos em como a moderação de conteúdo é feita e desencadear censura colateral. O responsável em potencial (plataforma ou administrador de grupo de mensagens), por receio em arcar com ônus de eventual conteúdo considerado infrator, teria incentivo para efetuar maior restrição ao conteúdo que pode ser compartilhado nesses espaços. Assim, este projeto de lei, se aprovado, pode impedir a livre formação de grupos de discussão e compartilhamento online.

Pode-se perceber, pela análise da regulação queniana, que grande parte das medidas regulatórias voltadas a conteúdo online não se traduz em obrigações de moderação de conteúdo, mas em punições e/ou restrições para sua circulação. O fato de haver escassas medidas de transparência na regulação reforça, ainda, o cenário de opacidade sobre as medidas de moderação de conteúdo. Isso também dificulta a abertura para discutir as restrições efetuadas, porque as medidas restritivas adotadas não são objeto de debate público.

3. Síntese comparativa

A comparação entre os cenários examinados permite identificar pontos de aproximação e discrepância entre os tratamentos regulatórios em uso para a atividade de moderação de conteúdo e parâmetros de transparência.

3.1. Tema e contexto das regulações

O principal consenso é a preocupação regulatória em torno do conteúdo considerado danoso, o que demonstra a percepção dos agentes reguladores sobre a insuficiência do modelo autorregulatório. Dentre os países analisados, a Alemanha destaca-se com a NetzDG, lei que obriga plataformas a tomarem providências quanto ao conteúdo, assim como a Índia, com as Diretrizes de Intermediários, cujos efeitos ainda estão sob análise. Nessa mesma linha, percebem-se projetos em trâmite avançado e que prevêem também maior responsabilidade das plataformas, como o Projeto de Lei de Segurança Online na Austrália, o Projeto de Lei 2630/2020 no Brasil e o Projeto de Emenda à Lei de Informação e Comunicações no Quênia.

Também é perceptível a conexão das discussões com o contexto político e eleitoral. Vale destacar que há um guia sobre políticas de cibersegurança em eleições da

Commonwealth[254], da qual Austrália, Índia e Quênia fazem parte. Neste documento consta a recomendação para que países obriguem plataformas e redes de propaganda a publicizar anúncios políticos, para permitir consulta e escrutínio público. Ele também determina que sejam feitos acordos com redes sociais a fim de aderir a um conjunto de práticas, a serem definidas com base no Código de Prática relativo à desinformação da União Europeia. Ainda, recomenda fortalecer a publicação de gastos com conteúdo político e doações realizadas para influenciar campanhas[255]. Assim, percebe-se que, na *Commonwealth*, há uma tônica de promover transparência nas plataformas aliada à maior democratização dos processos eleitorais e que isso passa por uma identificação mais fiel desse tipo de conteúdo e de suas fontes financeiras. Esse é o tema explícito da diretriz queniana da NCIC, que busca regular disseminação de propaganda política na internet em época de eleições.

A atenção voltada ao discurso de ódio na Alemanha, onde há partidos extremistas que propagam este tipo de conteúdo, e à desinformação na Índia e no Brasil, que são países nos quais os governantes e candidatos lançam mão de controle midiático, vincula-se também ao receio de que esses conteúdos prejudiquem o debate público. Entretanto, via de regra, essas propostas impõem obrigações de monitoramento ou resposta às plataformas, e isso possibilita que haja também excesso quanto às medidas adotadas para atender os prazos e evitar responsabilização.

O risco de regulações que focam em controle de conteúdo é a censura colateral sobre conteúdo legítimo, especialmente quando são legislações não específicas sobre um determinado contexto. Nesse sentido, o Brasil se destaca das demais jurisdições por adotar a autorregulação das plataformas na maioria dos casos, sem sua responsabilização pelo conteúdo de terceiros, por dispositivo no Marco Civil da Internet (embora haja um movimento de mudança, com as novas responsabilidades propostas no PL 2630/2020).

Outro ponto de convergência entre alguns países é a forte influência da proteção a direitos autorais na moderação de conteúdo. Essa matéria foi mais relevante na regulação da Alemanha, que conta com regime rígido de controle de conteúdo protegido, bem como na Austrália, onde um dos principais motivos para remoção de conteúdo é a legislação de direitos de autor.

De maneira geral, o poder das plataformas de moderar conteúdo é tratado como objeto de exigências contextuais de cada país, como uma solução a demandas ocasionadas pela volumosa produção, circulação e compartilhamento de conteúdo pelos usuários. Isso resulta em cenários regulatórios com mais demandas do que

254 COMMONWEALTH. **Cybersecurity for Elections**: a Commonwealth guide on best practice. Disponível em: https://books.thecommonwealth.org/cybersecurity-elections-paperback Acesso em 16 jun. 2021.

255 Ibidem, p. 136.

limitações a essa atividade, o que amplifica os efeitos desse controle[256] – seja ele feito por exigência governamental ou por autorregulação. A fim de que esse poder não seja exercido de maneira abusiva, um primeiro passo é que a regulação ofereça ferramentas de fiscalização sobre ele ao público afetado. Entretanto, como melhor examinado a seguir, esse ponto é bastante deficitário na amostragem deste estudo.

3.2. Papel da transparência na regulação analisada

De acordo com os requisitos objetivos de transparência na moderação de conteúdo recomendados por diversas organizações internacionais e discutidos anteriormente, percebe-se que nenhum país estabeleceu obrigações robustas de transparência. Seria esperado o estabelecimento de parâmetros obrigatórios de devido processo que permitissem a maior publicidade possível das decisões e critérios para restrição a conteúdo, tanto ao usuário responsável quanto ao público em geral – com equilíbrio em relação à proteção de dados pessoais dos usuários envolvidos.

Entre essas obrigações estaria a publicação de políticas que orientam a moderação, bem como os critérios para enquadramento da atividade. Ainda, a regulação poderia prever a notificação do usuário responsável pelo conteúdo, com exposição de motivo para a medida cogitada e as especificidades relativas às violações de política de comunidade. Outras previsões perdiam envolver a oportunização de mecanismo de revisão da decisão mediante novas informações, e à moderação decorrente de ordem de autoridade. A publicação de relatórios de transparência que permitia analisar os padrões seguidos na atividade de moderação na prática também é um dever relevante. Outro aspecto importante é o dever de transparência algorítmica, tendo em conta que essa atividade conta com amplo apoio de máquinas e que não há obrigação de informar quando isso é feito nem quais os meios de revisão dessas decisões. Embora algumas dessas práticas sejam adotadas e a literatura as relate, não existe uma padronização obrigatória e em muitos casos a transparência que elas proporcionam é opaca, apenas instrumental, e não garante a possibilidade de fiscalização.

Identificaram-se alguns casos em que as obrigações dirigiam-se a deveres das plataformas em relação ao conteúdo em si, como no projeto queniano e no brasileiro, ou na lei alemã.Esse tipo de regulação, contudo, é desencorajado, especialmente quando desacompanhado de obrigações de submissão a controle social dessas medidas.

256 Uma crítica concreta a regulações que ocasionam favorecimento institucional da restrição a conteúdos online, em detrimento das garantias à livre expressão, é feita em: KELLER, Daphne. TheRight Tools: Europe's Intermediary Liability Laws and the EU 2016 General Data Protection Regulation. **Berkeley Technology Law Journal**, v. 33, p. 287, 2018. Disponível em: https://btlj.org/data/articles2018/vol33/33_1/Keller_Web.pdf Acesso em: 27 jun. 2021.

Obrigações de moderação de conteúdo

Os países que estabelecem – ou buscam estabelecer, no caso do Brasil – obrigações de monitoramento ativo são aqueles que incorporam mais requisitos de transparência em seus ordenamentos jurídicos. Ainda assim, considerando os riscos do monitoramento proativo de conteúdo, essas obrigações são insuficientes na maioria das legislações exploradas. Este tipo de obrigação de monitoramento implica na legitimação, pelo Estado, de um constante estado de vigilância privada no qual empresas que já tratam altos volumes de dados pessoais estejam autorizadas a apropriarem-se de informações ainda mais íntimas de indivíduos e coletividades. Quanto mais poderes são delegados a entes privados, mais mecanismos de controle e acompanhamento do exercício de tais poderes devem ser estabelecidos de forma a mitigar os abusos e violações de direitos.

Dever de notificar

O dever de notificar a pessoa afetada pela intervenção da plataforma em determinado conteúdo de terceiro é uma obrigação presente em todos os ordenamentos que exigem o monitoramento ativo de conteúdo. Entretanto, o Quênia exige essa notificação apenas em casos em que a plataforma veta o envio em massa de mensagem política. Nenhum outro tipo de moderação de conteúdo além da remoção da publicação precisa ser informado. A notificação do usuário é a primeira etapa de um sistema de revisão de decisões resultantes da moderação de conteúdo e seu objetivo final é possibilitar que alguém questione a moderação e defenda a disponibilização do conteúdo. Assim, a legislação queniana, ao negligenciar a notificação de outros tipos de moderação, desconsidera a etapa mais básica para que eventuais práticas abusivas de interferência de conteúdo sejam contestadas e que conteúdos legítimos sejam revistos.

Apesar de também dispor sobre a obrigação de notificação, a legislação indiana é extremamente abrangente quanto ao usuário que deve ser comunicado. Ele pode ser aquele que “criou, carregou, compartilhou, disseminou ou modificou” o conteúdo. Tal definição genérica pode resultar em uma legislação inefetiva e, por conseguinte, em uma transparência opaca, ou seja, acentua obscuridades ao invés de aumentar a racionalidade. Apesar das recomendações internacionais também se referirem ao usuário afetado em sentido amplo, o estabelecimento dessa regra no âmbito de determinada jurisdição deve ocorrer com base na viabilidade técnica de seu cumprimento.

Ainda quanto à imprecisão dos dispositivos legais sobre notificação de usuário elencados nesta pesquisa, percebe-se que diversas informações que devem constar na notificação não são incorporadas como obrigatórias pelas legislações. Tão importante quanto saber que houve uma interferência em determinado conteúdo

é saber qual foi o tipo de interferência realizado, o que motivou a moderação, qual a descrição exata do conteúdo que foi afetado e como a plataforma detectou aquele conteúdo. Entre os países analisados, o Brasil possui uma proposta legislativa que obriga os intermediários a notificarem os usuários incluindo todas as informações necessárias. Entretanto, até mesmo este projeto de lei possui lacunas quanto à notificação, visto que dispensa o ato em algumas hipóteses, como o comprometimento da estabilidade da aplicação ou segurança do usuário.

Relatórios de transparência

Uma segunda obrigação bastante presente nas legislações analisadas é a exigência de publicação de relatórios de transparência pelas plataformas. As recomendações internacionais nesse quesito são robustas. Entre as informações que devem constar nos relatórios, destacam-se informações sobre restrição de conteúdo em razão de requisição governamental, ordem judicial, requisição de agentes privados ou aplicação das políticas da plataforma, além de ser necessário mencionar o tipo de conteúdo removido ou interferido, localização geográfica dos envolvidos e fonte da denúncia.

Obrigações robustas de transparência são encontradas no projeto de lei do Brasil e na legislação da Alemanha, enquanto que a regulação da Índia apresenta obrigações menos vultosas. No caso do Brasil, o PL 2630/20 demanda a publicação de relatórios trimestrais sobre as características dos setores de moderação das empresas e os números de medidas de moderação adotadas, bem como a metodologia, motivação e tipo de medida adotada. Entretanto, o projeto brasileiro em tramitação não exige alguns dos requisitos recomendados pelas organizações mencionadas neste estudo. Não existe a demanda de informar nos relatórios as requisições para restrição de conteúdo, tampouco a localização geográfica dos envolvidos.

Na Alemanha, o NetzDG determina que plataformas com mais de 100 denúncias sobre conteúdo ilegal ao ano elaborem relatórios de transparência com periodicidade semestral. Os requisitos mínimos para esses documentos incluem informações sobre os esforços empregados pelas plataformas para endereçar as denúncias, os mecanismos disponíveis para que os usuários possam reportar os abusos, o número de denúncias e o tempo necessário para apreciação de cada uma delas, as soluções empregadas para cada caso, entre outras.

Entretanto, a maior parte das legislações analisadas não detalham quais informações devem constar nos relatórios de transparência. Desta forma, é importante que os relatórios de transparência exigidos ofereçam o máximo de informações sobre suas práticas de moderação de forma agregada para que pesquisadores e formuladores de políticas públicas possam avaliar a atuação em nível global. Cabe, contudo, enfatizar que esses esforços devem ser compatibilizados com as normas e os princípios da privacidade dos usuários e da proteção de seus dados pessoais.

A obrigação de publicar relatório também é presente na legislação indiana. As Regras de Tecnologia da Informação apenas mencionam o dever dos intermediários de publicar mensalmente relatórios sobre reclamações e respostas à reclamações, bem como o número de conteúdos nos quais a plataforma interferiu (seja conteúdo removido ou desabilitado). Além de não incluir diversos requisitos recomendados internacionalmente quanto à publicação de relatórios, a norma indiana não é específica o suficiente sobre quais informações sobre as reclamações e atuação da plataforma devem ser oferecidas, o que pode resultar em transparência de fachada.

Considerando que as pesquisas e consultas à comunidade acadêmica e sociedade civil resultaram em uma lista de informações a serem publicizadas de forma ampla, é importante que as legislações especifiquem os dados a serem apresentados nos relatórios, sob pena de se tornarem inefetivas e violarem a proteção de dados pessoas. Além disso, ao encontro da crítica apresentada neste estudo sobre relatórios que não contextualizam os dados apresentados e que podem gerar desinformação, não foi encontrada em nenhuma legislação uma preocupação com a interpretação e forma de disponibilização dos dados. Esse ponto também não é endereçado pelas recomendações internacionais que baseiam a análise deste trabalho. No entanto, é importante destacar que legislações futuras devem oferecer estruturas para que os relatórios de transparência sejam inteligíveis e assertivos.

Obrigações em caso de restrição de conteúdo

O terceiro ponto a ser destacado na comparação diz respeito às obrigações em casos de restrições de conteúdo. Como já mencionado, a Índia as impõe e o Quênia também propõe em um projeto o monitoramento proativo de conteúdo e a consequente interferência quando necessária – o que frequentemente é a exclusão do conteúdo. Além dos países que exigem tal moderação, existem aqueles que a autorizam de forma voluntária e de acordo com os próprios interesses e regras das plataformas. Em ambos os casos, as recomendações internacionais apontam para a necessidade de que qualquer interferência no conteúdo de terceiro seja feita de acordo com determinados critérios e procedimentos.

Um dos requisitos prévios à moderação de conteúdo é a explicitação, na política de comunidade, dos comportamentos proibidos na plataforma. Este requisito guarda relação com o princípio da legalidade na medida em que existe para garantir que ninguém tenha sua manifestação cerceada em razão de regra desconhecida. Entretanto, apenas a legislação indiana obriga os intermediários a informar periodicamente os usuários acerca das regras que moldam as permissões de discurso na plataforma. No Brasil, existe uma proposta legislativa que menciona vagamente a necessidade de tais regras constarem nos termos de uso da plataforma e endereça a responsabilidade de fiscalizar e acompanhar a obrigação a um Conselho Multissetorial a ser criado pelo Congresso Nacional. Quênia e Austrália

não estabelecem nenhuma obrigação quanto às políticas de comunidade.

Em estudo anterior[257], o IRIS analisou diversas políticas de comunidade de redes sociais com o intuito de avaliar o grau de transparência e a forma como os documentos endereçaram a questão. A partir das recomendações desse estudo, percebe-se uma ausência de preocupação dos Estados analisados com as políticas de comunidade, que muitas vezes são o fundamento da interferência em diversos conteúdos digitais.

Transparência algorítmica

Apesar de ser frequente o estímulo ao uso de ferramentas de automatização de decisões, como ocorre na Índia, as obrigações de transparência algorítmica são escassas. O Brasil, a Alemanha e a Austrália estipulam regras sobre o tema.

No caso do Brasil, existem propostas que estabelecem direitos aos interessados nas decisões provenientes de uso de inteligência artificial, como o acesso à informações nítidas sobre os critérios e procedimentos utilizados pelos sistemas de inteligência artificial. Entretanto, a legislação é extremamente abrangente em relação a estes critérios e não define outras obrigações em relação à decisão automatizada, como a publicação de relatórios periódicos e auditoria dos sistemas.

Já a Austrália conta com enunciados que determinam a prestação de informações sobre como funcionam os algoritmos de ranqueamento de conteúdo online. Essa obrigação, contudo, aplica-se apenas a plataformas de veiculação de conteúdo gerado por terceiros. Já a obrigação de transparência é devida apenas a veículos jornalísticos que disponibilizam seu conteúdo através dessas plataformas – os usuários regulares, portanto, não são abrangidos na legislação australiana.

A Alemanha, por sua vez, atribui às plataformas a obrigação de transparência quanto ao funcionamento de seus sistemas de ranqueamento de conteúdo. Essa obrigação é direcionada a prestadoras de serviços de naturezas diversas, e – diferentemente das determinações australianas sobre funcionamento algorítmico – se estende também à prestação de informações aos próprios usuários desses serviços. Para isso, prevê o fornecimento de informações nítidas e facilmente acessíveis sobre o funcionamento dos algoritmos das plataformas. Adicionalmente, a proposta de alteração ao NetzDG ainda em tramitação durante a elaboração deste estudo prevê a inserção de informações relativas à lógica e tomada de decisão envolvidas na moderação de conteúdo aos relatórios de transparência já mencionados no texto legal.

Pode-se perceber que são esparsas as iniciativas voltadas à transparência na moderação de conteúdo na amostra analisada, como mostra o quadro-síntese abaixo:

257 RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021

Quadro 1 - Síntese comparativa da regulação sobre transparência na moderação de conteúdo na amostra analisada

	ALEMANHA	AUSTRÁLIA	BRASIL	ÍNDIA	QUÊNIA
Obrigações de moderação de conteúdo	Determinação para que as plataformas proativamente realizem o bloqueio ou remoção de conteúdo ilícito que for denunciado, sob pena de multa; obrigação de repassar a autoridades dados de pessoas que compartilharem conteúdo contendo discurso de ódio.	Conteúdo considerado "abominavelmente violento" que for facilmente acessível pelo público australiano deve ser prontamente bloqueado ou removido das plataformas, sob pena de multa ou até mesmo responsabilização penal	Regras específicas relativas à remoção de conteúdo sexual ou nudez; Proposta: dever de regras de restrição constarem nos termos de uso da plataforma; fiscalização e acompanhamento da obrigação por Conselho Multissetorial; monitoramento ativo de conteúdo e contas.	Intermediários devem informar periodicamente os usuários sobre políticas de comunidade; proibição de alguns tipos de conteúdo; obrigação de monitoramento proativo de conteúdo e contas.	Mensagem política em massa enviada por meio de acordo entre aplicativos de conteúdo e operadoras de rede móvel devem ter seu conteúdo submetido previamente às últimas, as quais, em caso de dúvida, devem enviá-lo para análise de adequação da autoridade.
Dever de notificar	Decisões relativas a denúncias sobre conteúdo ilícito devem ser notificadas e justificadas tanto para a pessoa que publicou o material quanto para a que o denunciou.	-	Proposta: obriga intermediários a notificarem os usuários incluindo todas as informações necessárias, mas dispensa quando a estabilidade da aplicação ou a segurança do usuário estiver em risco.	Dever de comunicar o usuário que "criou, carregou, compartilhou, disseminou ou modificou" o conteúdo a ser moderado.	Exigência de notificação apenas em casos em que a plataforma veta o envio em massa de mensagem política.

	ALEMANHA	AUSTRÁLIA	BRASIL	ÍNDIA	QUÊNIA
Relatórios de transparência	Plataformas que recebem mais de 100 denúncias sobre conteúdo ilícito ao ano devem elaborar relatórios de transparência semestrais e publicar de forma detalhada e segmentada como são endereçadas as referidas denúncias.	Proposta: a autoridade pode requerer publicação de relatório sobre cumprimento de medidas de segurança digital.	Proposta: publicação de relatórios trimestrais sobre as características dos setores de moderação das empresas e os números de medidas de moderação adotadas, bem como a metodologia, motivação e tipo de medida adotada.	Dever dos intermediários de publicar mensalmente relatórios sobre reclamações e respostas a reclamações, bem como o número de conteúdos nos quais a plataforma interferiu (seja conteúdo removido ou desabilitado).	-
Obrigações em caso de restrição de conteúdo	-	-	Proposta: direito de recorrer e direito de resposta.	Direito de recorrer, mecanismos de resolução de disputas supervisionado.	Notificação em caso de veto a envio de mensagem política em massa.
Transparência algorítmica	Obrigação de transparência quanto ao funcionamento de seus sistemas de ranqueamento de conteúdo.	Obrigação de transparência quanto ao funcionamento de seus sistemas de ranqueamento de conteúdo, porém as informações são devidas apenas a veículos de mídia jornalística que disponibilizam seus conteúdos através das plataformas.	Proposta: acesso a informações nítidas sobre os critérios e procedimentos utilizados pelos sistemas de inteligência artificial.	-	-

Fonte: autoria própria

4. Conclusão

Esta pesquisa evidenciou o quanto é relevante que práticas de moderação de conteúdo na internet sejam transparentes. A conceituação dessa atividade inclui algumas práticas de intervenção em conteúdo gerado por usuário e também as implicações de que elas sejam realizadas por meio de trabalho humano ou automatizado por inteligência artificial. Pôde-se perceber a escala e o potencial efeito que esse controle tem sobre a liberdade de expressão, direito humano que orienta também a operação da internet como tecnologia. Esses fatores indicam a necessidade de instrumentos para os usuários conhecerem os critérios, meios e técnicas que baseiam a atuação das plataformas.

Nesse sentido, a emergência de uma discussão regulatória sobre as obrigações das plataformas em relação ao conteúdo torna-se interessante objeto de estudo. Por meio da análise de documentos internacionais considerados como referências para uma atuação baseada em direitos humanos e nos princípios que sustentam uma harmonização global de instituições e práticas, observam-se algumas tendências. Uma delas é a de apontar que a regulação sobre intermediários considere que obrigações impostas a eles podem ter efeitos inibitórios sobre direitos dos usuários. Nessa linha, recomendam-se regulações e adoção de práticas que permitam o escrutínio público de intervenções que intermediários façam sobre o conteúdo de terceiros. Ainda, há desencorajamento à regulação restritiva ao conteúdo que circula online, seja pela criminalização de usuários ou pela responsabilização dos intermediários em monitorá-lo proativamente. Uma série de recomendações sobre transparência pode ser identificada em instrumentos principiológicos e relatórios de especialistas, o que gera a expectativa de que sejam incorporadas em regulações locais.

Já na análise de cenários regulatórios nacionais em diferentes contextos, percebe-se nesta seleção a crescente expectativa de maior exercício do poder de intervenção sobre conteúdo. Em diversos âmbitos regulatórios, os problemas que se apresentam pelo fluxo de informações produzidas e circuladas por usuários – polarização, desinformação, discurso de ódio, violação de direitos autorais – são alvo de regulações que buscam sua solução pelo controle à expressão. Novos deveres e responsabilidades, bem como novos mecanismos – tais como inteligência artificial – são incentivados como resposta aos conflitos provenientes da comunicação em rede. Essas medidas, via de regra, consolidam e amplificam os poderes de agentes privados sobre direitos como livre expressão, acesso à informação e autodeterminação informativa.

Ademais, fica à margem dessas regulações impor, concretamente, medidas que permitam fiscalizar esse poder. Há uma lacuna quanto às obrigações que envolvem transparência na moderação de conteúdo. A amostra de regulações examinadas

permite concluir que não há um conjunto padrão de obrigações para explicitar de que forma o conteúdo é restrito e como serviços que ofertam comunicação estabelecem medidas para limitá-la. Isto é, há uma gama de instrumentos regulatórios e normativos referentes às atividades de plataformas de conteúdo gerado por usuário, com enfoques distintos em cada jurisdição, porém com um ponto em comum: há uma preocupação com mais regulação sobre a atividade de moderação de conteúdo, mas não focada em tornar essas atividades passíveis de escrutínio pelo usuário ou mesmo por entidades de fiscalização.

Mesmo nas jurisdições que não adotam autorregulação, em que conteúdos específicos são regulados por lei e as plataformas estariam adstritas a essas regras para algumas atividades de moderação, esse seria um ponto essencial para a certificação de que abusos não estão sendo cometidos frente aos poderes confiados às plataformas. Sua ausência é especialmente preocupante tendo em conta o levantamento de conteúdo midiático nos países pesquisados, que demonstra o interesse de atores políticos extremistas em controlar meios de comunicação e mídias. Os problemas que a moderação resolveria, nesse cenário, podem acabar fortalecidos a depender do controle que pode ser exercido extra institucionalmente sobre as plataformas. Ainda, uma sobrecarga de deveres pode levar plataformas a exercerem de forma abusiva a atividade de controle de conteúdo, o que afeta também conteúdo legítimo, relevante e necessário.

Sugere-se, tendo em conta a relevância do assunto e também a contínua transformação dos dispositivos e das políticas regulatórias, que novos estudos sejam realizados sobre a temática. Isso tanto para averiguar os efeitos e desdobramentos das leis e projetos em trâmite nos países aqui descritos, quanto para ampliar o escopo da pesquisa. Estes resultados incluíram atualizações até abril de 2021 dentro das jurisdições analisadas. Tampouco foi possível retratar tendências regulatórias regionais devido à limitação da amostra, que prezou pela diversidade de regiões abrangidas. Isso implicou em análise de apenas um país de cada continente, com a limitação linguística ao inglês, português e espanhol. Entende-se que pode haver variações significativas dentro de cada região e aspectos que também podem ser identificados como tendências daquele continente específico caso sejam realizados estudos com enfoque em vários países de uma mesma região.

Feitos esses apontamentos, entende-se que esta pesquisa contribui com o conhecimento do estado da arte da regulação sobre um tema emergente envolvendo direitos na internet, que é a moderação de conteúdo, em uma perspectiva comparada e crítica. O estudo pode ser complementado com uma pesquisa futura sobre quais os modelos de responsabilidade sobre conteúdo de terceiros que as regulações inspiram e quais os efeitos de cada um na prática. Assim, espera-se que sirva de base para novas discussões acerca dos desdobramentos desses dispositivos legais, bem como estudos sobre a forma como eles interagem com a maneira de ofertar serviços e, sobretudo, com os direitos dos usuários de internet.

Referências

ABRAJI - Associação Brasileira de Jornalismo Investigativo. **Políticos tentaram ocultar informação mais de 500 vezes durante a campanha**. 21 nov. 2018. Disponível em: https://www.ctrlx.org.br/noticia/politicos-tentaram-ocultar-informacao-mais-de-500-vezes-durante-a-campanha Acesso em 19 maio 2021.

ABRIL. Tribunal do Quênia mantém criminalização da homossexualidade. **Veja**, 24 maio 2019. Disponível em: https://veja.abril.com.br/mundo/tribunal-do-quenia-mantem-criminalizacao-do-homossexualismo/ Acesso em: 14 jun. 2021.

AGARWAL, Surabhi. Officials debate whether India should have its own social media content moderation rules. **The Economic Times**. 1 set. 2020. Disponível em: https://news.google.com/articles/CAIiELSW6qmuRu6lx_AX0NtvBJkqGQgEKhAIACoHCAow2pqGCzD954MDMMLNigY?hl=en-IN&gl=IN&ceid=IN%3Aen. Acesso em 19 maio 2021.

AGUIAR, L., CLAUSSEN, J., PEUKERT, C. . Catch Me if You Can: Effectiveness and Consequences of Online Copyright Enforcement. **Information Systems Research**, 29(3), p. 656–678. 2018. Disponível em: https://research-api.cbs.dk/ws/portalfiles/portal/57343396/j_rg_claussen_et_al_catch_me_if_you_can_acceptedversion.pdf. Acesso em: 19 maio 2021.

ALEMANHA. BUNDESLÄNDER. **Tratado de Mídia Estatal**. Berlim, 28 de abril de 2020. Disponível em: https://www.gesetze-bayern.de/Content/Document/MStV/true. Acesso em 17 jun. 2021.

ALEMANHA. BUNDESLÄNDER. **Tratado Interestadual de Transmissão**. Berlim, 31 de agosto de 1991. Disponível em: https://recht.nrw.de/lmi/owa/br_bes_text?anw_nr=2&gld_nr=2&ugl_nr=2251&bes_id=12784&aufgehoben=N&menu=1&sg=0. Acesso em 17 jun. 2021.

ALEMANHA. BUNDESTAG. Código Criminal. Berlim, 15 de maio de 1871. Disponível em: https://www.gesetze-im-internet.de/englisch_stgb/. Acesso em 17 jun. 2021.

ALEMANHA. BUNDESTAG. **Lei de Direitos Autorais e Direitos Correlatos**. Berlim, 09 de setembro de 1965. Disponível em: https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html. Acesso em 17 jun. 2021.

ALEMANHA. BUNDESTAG. **Lei de Regulação da Internet**. Berlim, 01 de setembro de 2017. Disponível em: https://germanlawarchive.iuscomp.org/?p=1245. Acesso em 17 jun. 2021.

ALEMANHA. BUNDESTAG. **Lei de Telecomunicações**. Berlim, 22 de junho de 2004. Disponível em: https://germanlawarchive.iuscomp.org/?p=692. Acesso em 17 jun. 2021.

ALEMANHA. BUNDESTAG. **Lei para Combate ao Extremismo de Direita e ao Crime de Ódio**. Berlim, 30 de março de 2021. Disponível em: https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/Bgbl_Bekaempfung_Rechtsextremismus_Hasskriminalitaet.pdf?__blob=publicationFile&v=3. Acesso em 17 jun. 2021.

ALGORITHM WATCH. **Germany's new media treaty demands that platforms explain algorithms and stop discriminating**. Can it deliver? Disponível em: https://algorithmwatch.org/en/new-media-treaty-germany/. Acesso em 19 maio 2021.

ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New media & society**, v. 20, n. 3, p. 973-989, 2018. Disponível em: https://journals.sagepub.com/doi/10.1177/1461444816676645 Acesso em: 18 jun. 2021.

ANJOS, Lucas. **Testando os limites do capitalismo de vigilância**. 2019. Disponível em: https://irisbh.com.br/testando-os-limites-do-capitalismo-de-vigilancia/. Acesso em: 14 jun. 2021.

ANTUNES, Laila Damascena; ROSA, Matheus; VILELA, Pedro. **Jurisdição e internet:** Estudo sobre mecanismos de bloqueio e fragmentação da rede. Instituto de Referência em Internet e Sociedade: Belo Horizonte, 2017. Disponível em: http://bit.ly/2PjUqYT. Acesso em: 14 jun. 2021.

AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Emenda ao Código Penal** (Compartilhamento de Material Abominavelmente Violento). Camberra, 05 de abril de 2019. Disponível em: https://www.legislation.gov.au/Details/C2019A00038. Acesso em: 17 jun. 2021.

AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei de Direitos Autorais**. Camberra, 01 de maio de 1969. Disponível em: https://www.legislation.gov.au/Details/C2021C00044. Acesso em: 17 jun. 2021.

AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei de Emenda às Leis do Tesouro** (Código de Negociação Obrigatório para Plataformas Digitais e Mídia de Notícias). Camberra, 09 de dezembro de 2020. Disponível em: https://www.legislation.gov.au/Details/C2020B00190. Acesso em: 17 jun. 2021.

AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei para Emenda de Legislação**

Relativa a Telecomunicações e Outros (Assistência e Acesso). Camberra, 08 de dezembro de 2008. Disponível em: https://www.legislation.gov.au/Details/C2018A00148. Acesso em: 17 jun. 2021.

AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Lei Para Melhorar a Segurança Online**. Camberra, 10 de setembro de 2018. Disponível em: https://www.legislation.gov.au/Details/C2018C00356. Acesso em: 17 jun. 2021.

AUSTRÁLIA. HOUSE OF REPRESENTATIVES. **Projeto de Lei para Segurança Online**. Camberra, 24 de fevereiro de 2021. Disponível em: https://www.legislation.gov.au/Details/C2021B00018. Acesso em: 17 jun. 2021.

AVAST. **Doxxing**: O que é e como evitá-lo. Disponível em: https://www.avast.com/pt-br/c-what-is-doxxing. Acesso em: 17 jun. 2021.

BALKIN, Jack M. Free speech is a triangle. **Columbia Law Review**., v. 118, p. 2011, 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3186205 Acesso em: 18 jun. 2021.

BALKIN, Jack M., Digital Speech and Democratic Culture: a Theory of Freedom of Expression for the Information Society. **New York University Law Review**, Vol. 79, No. 1, 2004, Yale Law School, Public Law Working Paper No. 63, Disponível em: https://ssrn.com/abstract=470842 Acesso em 11 maio 2021.

BARTLE, Ian; VASS, **Peter. Self-regulation and the regulatory state**: A survey of policy and practice. Bath: Centre for the Study of Regulated Industries, University of Bath School of Management, 2005. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.3190&rep=rep1&type=pdf. Acesso em 15 jul. 2021.

BBC. **Facebook blocks Australian users from viewing or sharing news**. 18 fev. 2021. Disponível em: https://www.bbc.com/news/world-australia-56099523. Acesso em: 19 maio 2021.

BENTES, Ana. A gestão algorítmica da atenção: Enganchar, Conhecer e Persuadir In: **Seminário Governança das Redes**, 4., 2019, Belo Horizonte. Políticas, Internet e Sociedade. Belo Horizonte: IRIS. 984. p. 222-234. Disponível em: https://irisbh.com.br/publicacoes/politicas-internet-e-sociedade/. Acesso em: 13 maio 2021.

BIDDLE, Sam; RIBEIRO, Paulo Victor; DIAS, Tatiana. Censura invisível. Exclusivo: TikTok censurou rostos feios e favelas para atrair novos usuários e censurou posts políticos. **The Intercept Brasil**. 16 mar. 2020. Disponível em: https://theintercept.com/2020/03/16/tiktok-censurou-rostos-feios-e-favelas-para-atrair-novos-usuarios/ Acesso em: 12 fev. 2021.

BIRCHALL, Clare. Introduction to 'Secrecy and Transparency' The Politics of Opacity and Openness. **Theory, Culture & Society**, v. 28, n. 7-8, p. 7-25, 2011. Disponível em: https://journals.sagepub.com/doi/10.1177/0263276411427744. Acesso em 14 jun. 2021.

BITKOM. **Ancillary Copyright for Publishers: taking stock in germany**. Berlim: Bitkom, 2015. Disponível em: https://www.bitkom.org/sites/default/files/file/import/160406-Leistungsschutzrecht-EN1.pdf. Acesso em: 19 maio 2021.

BLACK, Julia. Transparent policy measures. **Oxford Dictionary of Economics**. Oxford University Press, Oxford, 1997. Disponível em: https://www.oxfordreference.com/view/10.1093/acref/9780199237043.001.0001/acref-9780199237043-e-3191. Acesso em 13 maio 2021.

BOWERS, John; ZITTRAIN, Jonathan L. Answering impossible questions: Content governance in an age of disinformation. **The Harvard Kennedy School (HKS) Misinformation Review.** 10 fev. 2020. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3520683 Acesso em 15 jun. 2021.

BRASIL, CÂMARA DOS DEPUTADOS. **Projeto de Lei nº 2.630, 03 jul. 2020.** (Do Sr. Alessandro Vieira). Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=node01099cclmqrw1y1s9ebm6006wp43046552.node0?codteor=1909983&filename=PL+2630/2020. Acesso em: 19 maio 2021.

BRASIL, CÂMARA DOS DEPUTADOS**. Projeto de Lei nº 21, 04 fev. 2020**. (Do Sr. Eduardo Bismarck). Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928&filename=PL+21/2020 Acesso em: 19 maio 2021.

BRASIL. Lei Geral de Proteção de Dados Pessoais (LGPD). Lei 13.709, de 14 de agosto de 2018. **Diário Oficial da União**, Brasília, DF, Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 20 maio 2021

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. **Diário Oficial da União**. Brasília, 24 abr. 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm Acesso em 19 maio 2021.

BRASIL. Lei nº 13709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**. Brasília, DF, 14 ago. 2018. Dísponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm Acesso em 11 maio 2021

BRASIL. Ministério da Ciência, Tecnologia e Inovações. **Estratégia Brasileira de Inteligência Artificial é publicada no DOU**. 9 abr. 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2021/04/estrategia-brasileira-de-inteligencia-artificial-e-publicada-no-dou#:~:text=A%20Estrat%C3%A9gia%20Brasileira%20de%20Intelig%C3%AAncia%20Artificial%20tem%20como%20objetivos%3A%20contribuir,e%20formar%20profissionais%20para%20o Acesso em: 19 maio 2021.

BRASIL. STF - Supremo Tribunal Federal. **STF discutirá validade de regra do Marco Civil da Internet sobre responsabilização de sites e redes sociais**. 2 mar. 2018. Disponível em: http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=371229&ori=1 Acesso em 19 maio 2021.

CALIXTO, Dodô. Mídia internacional destaca pioneirismo do Brasil com Marco Civil da internet. **Opera Mundi.** 26 mar 2014. Disponível em: https://operamundi.uol.com.br/politica-e-economia/34544/midia-internacional-destaca-pioneirismo-do-brasil-com-marco-civil-da-internet Acesso em 19 maio 2021.

CAMBRIDGE CONSULTANTS. **Use of AI in online content moderation**. 2019. Disponível em: https://www.epra.org/news_items/ofcom-s-report-the-use-of-ai-in-content-moderation. Acesso em 13 maio 2021.

CAPLAN R.; GILLESPIE, T. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. **Social Media + Society**. Jun. 2020. Disponível em: https://journals.sagepub.com/doi/full/10.1177/2056305120936636#articleCitationDownloadContainer. Acesso em: 3 fev. 2021..

CARVALHO, Caio. Após acusações de racismo, Twitter promete atualizar algoritmo de corte de imagens. **Gizmodo Brasil**. 2 out. 2020. Disponível em: https://gizmodo.uol.com.br/twitter-atualizar-algoritmo-corte-imagens-racismo/ Acesso em: 12 fev. 2021.

CE - Conselho da Europa. **Guia dos Direitos Humanos para os Utilizadores de Internet.** 16 abr. 2014. Disponível em: http://www.odionao.com.pt/media/5185/GuiaDireitosHumanosUtilizadoresInternet.pdf Acesso em 11 maio 2021.

CE - Conselho da Europa. **Recommendation CM/Rec(2012)4 of the Committee of Ministers to member States on the protection of human rights with regard to social networking services.** 4 abr. 2012. Disponível em: https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b Acesso em: 11 maio 2021.

CGI - Comitê Gestor da Internet no Brasil. **Sobre**. Disponível em: https://www.cgi.br/sobre/. Acesso em 19 maio 2021.

CHRISTCHURCH Call: **to eliminate terrorist & violent extremist content online**. Disponível em: https://www.christchurchcall.com/index.html. Acesso em: 19 maio 2021.

CHUNDURU, Aditya. Need Transparency In Facebook's Content Moderation Enforcement, Witnesses Tell Delhi's Peace And Harmony Committee. **Medianama**. 14 out. 2020. Disponível em: https://www.medianama.com/2020/10/223-need-transparency-facebook-content-moderation-delhi-peace-harmony-commiittee/. Acesso em: 10 maio 2021.

CIPESA - Collaboration on International ICT Policy for East and Southern Africa. **State of Internet Freedom in Kenya 2019**: mapping Trends in Government Internet Controls, 1999-2019. jan 2020. p. 20. Disponível em: https://www.opennetafrica.org/?wpfb_dl=100. Acesso em: 10 maio 2021.

CLARK, Justin et al. Content and Conduct: How English Wikipedia Moderates Harmful Speech. **The Social Science Research Network Electronic Paper Collection**, Research Publication, n. 2019-5, 2019. Disponível em: https://dash.harvard.edu/bitstream/handle/1/41872342/Content%20and%20ConductHow%20English%20Wikipedia%20Moderates%20Harmful%20Speech.pdf?sequence=1&isAllowed=y. Acesso em: 09 fev. 2021.

CLAYTON, James. Por que fundador do Twitter diz que banimento de Trump foi 'correto, mas perigoso'. **BBC**, 15 jan. 2021. Disponível em: https://www.bbc.com/portuguese/internacional-55674897 Acesso em: 11 maio 2021.

COMMONWEALTH. **Cybersecurity for Elections**: a Commonwealth guide on best practice. Disponível em: https://books.thecommonwealth.org/cybersecurity-elections-paperback Acesso em 16 jun. 2021.

COOKE, Jennifer. Background on the Post-Election Crisis in Kenya. **Blog CSIS** - Center for Strategic & International Studies. 6 ago. 2009. Disponível em: https://www.csis.org/blogs/smart-global-health/background-post-election-crisis-kenya Acesso em: 14 jul. 2021.

COOPER, Mariah. ARTS & ENTERTAINMENT Kenya lifts ban on lesbian film 'Rafiki,' giving it Oscar eligibility. **Washington Blade**. 21 set. 2018. Disponível em: https://www.washingtonblade.com/2018/09/21/kenya-lifts-ban-on-lesbian-film-rafiki-for-oscar-consideration/. Acesso em: 07 maio 2021.

DAQUINO, Fernando. Brasil: um dos maiores consumidores de eletrônicos. **Tecmundo**, 15 abr. 2013. Disponível em: https://www.tecmundo.com.br/mercado/38519-brasil-um-dos-maiores-consumidores-de-eletronicos.htm Acesso em: 19 maio 2021.

DEUTSCHE WELLE. Germany to abolish provider liability law, open path to more free WiFi. **Deutsche Welle.** Disponível em: https://www.dw.com/en/germany-to-abolish-provider-liability-law-open-path-to-more-free-wifi/a-19249024. Acesso em: 19 maio 2021.

EFF - Electronic Frontier Foundation et al. **Princípios de Manila sobre responsabilidade civil de intermediários**. Disponível em: https://manilaprinciples.org/pt-br/principles.html Acesso em: 14 jun. 2021.

EFF - Electronic Frontier Foundation et al. **Santa Clara Principles on transparency and accountability in content moderation**. Disponível em: https://santaclaraprinciples.org/ Acesso em: 14 jun. 2021.

ESTADÃO. Facebook pagará US$ 52 mi para funcionários com problemas na saúde mental. **Exame.** 14 maio 2020. Disponível em: https://exame.com/negocios/facebook-pagara-us-52-mi-para-funcionarios-com-problemas-na-saude-mental/. Acesso em 14 jun. 2021.

EU INTERNET REFERRAL UNIT. **Transparency Report**. 2017. Disponível em: https://www.europol.europa.eu/publications-documents/eu-internet-referral-unit-transparency-report-2017. Acesso em: 14 jun. 2021

EUROPEAN CONSUMER CENTRE GERMANY. **Illegal downloads in Germany**. Disponível em: https://www.evz.de/en/shopping-internet/illegal-downloads-in-germany.html. Acesso em: 19 maio 2021.

EUROPEAN UNION. **Code of conduct on countering illegal hate speech online**. 2016. Disponível em: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct. Acesso em: 21 jun. 2021.

FACEBOOK. **Content Restrictions Based on Local Law**. Disponível em: https://transparency.facebook.com/content-restrictions Acesso em 11 maio 2021.

FAGAN, Frank. Systemic social media regulation. **Duke Law & Technology Review**, v. 16, p. 393, 2017. Disponível em: https://scholarship.law.duke.edu/dltr/vol16/iss1/14/ Acesso em: 18 jun. 2021.

FERRI, F. The dark side(s) of the EU Directive on copyright and related rights in the Digital Single Market. **China-EU Law Journal** 2020. Disponível em: https://doi.org/10.1007/s12689-020-00089-5. Acesso em: 19 maio 2021.

FINEL, Bernard I.; LORD, Kristin M. 1999. The Surprising Logic of Transparency. **International Studies Quarterly** 43, no. 2, p. 316. Disponível em: https://transparencylab.org/Documentation/Additional%20resources/Literature/Finel%20et%20al_The%20Surprising%20Logic%20of%20Transparency%20_1999.pdf. Acesso em 06 mar. 2021.

FLYVERBOM, Mikkel. Digital age| transparency: Mediation and the management of visibilities. **International Journal of Communication**, v. 10, p. 13, 2016. Disponível em: https://ijoc.org/index.php/ijoc/article/view/4490 Acesso em: 18 jun. 2021.

FREEDOM HOUSE. **Brazil:** Freedom on the Net 2020. Disponível em: https://freedomhouse.org/country/brazil/freedom-net/2020 Acesso em 7 maio 2021.

FREEDOM HOUSE. **Kenya:** Freedom on the Net 2020. Disponível em: https://freedomhouse.org/country/kenya/freedom-net/2020 Acesso em 7 maio 2021.

FREEDOM ON THE NET. **Map:** internet freedom status. Disponível em: https://freedomhouse.org/explore-the-map?type=fotn&year=2019 Acesso em: 11 maio 2021.

FREITAS, Ariel. Mecanismo de busca do Google reforça o branco como padrão de beleza. **Notícia Preta**. 23 dez. 2019. Disponível em: https://noticiapreta.com.br/mecanismo-de-busca-do-google-reforca-o-branco-como-padrao-de-beleza/ Acesso em 11 maio 2021.

G1. **Twitter e Facebook bloquearam páginas e postagens nesta campanha eleitoral nos EUA;** relembre casos. 15 out. 2020. Disponível em: https://g1.globo.com/mundo/eleicoes-nos-eua/2020/noticia/2020/10/15/twitter-e-facebook-bloquearam-paginas-e-postagens-nesta-campanha-eleitoral-nos-eua-relembre-casos.ghtml Acesso em: 18 jun. 2021.

GILLESPIE, Tarleton L. et al. Expanding the Debate About Content Moderation.v. 9 n. 4, **Internet Policy Review**, 21 out. 2020

GILLESPIE, Tarleton. **Custodians of the Internet:** Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press, 2018. p. 217, p. 18

GILLESPIE, Tarleton. Platforms are not intermediaries. **Georgetown Law Technology Review**, v. 2, n. 2, p. 198-216, 2018. Disponível em: https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Gilespie-pp-198-216.pdf Acesso em: 18 jun. 2021.

GLOBAL NETWORK INITIATIVE. **Proposed German Legislation Threatens Free Expression Around the World.** 20 de abril de 2017. Disponível em: https://globalnetworkinitiative.org/proposed-german-legislation-threatens-free-expression-around-the-world/. Acesso em: 17 jun. 2021.

GNI - Global Network Initiative. **The GNI Principles Principles on Freedom of Expression and Privacy**. Disponível em: https://globalnetworkinitiative.org/gni-principles/ Acesso em: 14 jun. 2021.

GOLDMAN, Alex; VOGT, P.J. **Reply All**. Account Suspended. Disponível em: https://gimletmedia.com/shows/reply-all/94h3el6/171-account-suspended Acesso em: 03 fev. 2020.

GOMES, Alessandra; ANTONIALLI, Dennys; OLIVA, Thiago Dias. Drag queens e Inteligência Artificial: computadores devem decidir o que é 'tóxico' na internet? **InternetLab.** 28 jun. 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 11 maio 2021.

GOODFELLOW, Jessica. Pakistan reverses TikTok ban. **Campaign Asia,** 20 out. 2020. Disponível em: https://www.campaignasia.com/article/pakistan-reverses-tiktok-ban/464364. Acesso em 19 maio 2021.

GOOGLE. **Como funciona o Content ID.** Disponível em: https://support.google.com/youtube/answer/2797370?hl=pt-BR#zippy=%2Cquais-opções-estão-dispon%C3%ADveis-aos-proprietários-de-direitos-autorais. Acesso em: 04 fev. 2021.

GOOGLE. **Mel Silva's opening statement to the Senate Economics Committee Inquiry.** Disponível em: https://blog.google/around-the-globe/google-asia/australia/mel-silvas-opening-statement. Acesso em: 19 maio 2021.

GOOGLE. **Supporting Australian journalism: a constructive path forward** An update on the News Media Bargaining Code. Disponível em: https://about.google/google-in-australia/an-open-letter/. Acesso em: 19 maio 2021.

GOOGLE. **Update on the News Media Bargaining Code and Google in Australia**. Disponível em: https://about.google/google-in-australia/jan-6-letter/. Acesso em: 19 maio 2021.

HARRIS, Brent. **Establishing Structure and Governance for an Independent Oversight Board.** Facebook. 17 set. 2019. Disponível em: https://about.fb.com/news/2019/09/oversight-board-structure/. Acesso em 19 maio 2021.

HELDT, Amélie Pia. Reading between the lines and the numbers: an analysis of the first NetzDG reports. **Internet Policy Review**, v. 8, n. 2, 2019. Disponível em: https://policyreview.info/articles/analysis/reading-between-lines-and-numbers-analysis-first-netzdg-reports Acesso em: 18 jun. 2021.

HORWITZ, Jeff; PURNELL, Newley. ndia Threatens Jail for Facebook, WhatsApp and Twitter Employees. **The Wall Street Journal**, 5 mar. 2021. Disponível em: https://www.wsj.com/articles/india-threatens-jail-for-facebook-whatsapp-and-twitter-employees-11614964542. Acesso em 19 maio 2021.

IFF - Internet Freedom Foundation. **Deep dive:** How the intermediaries rules are anti-democratic and unconstitutional. 21 fev. 2021. Disponível em: https://internetfreedom.in/intermediaries-rules-2021/. Acesso em: 19 maio 2021.

INDIA. MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY. **Regras de Tecnologia da Informação** – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021. Nova Delhi, 25 de Fevereiro de 2021. Disponível em: http://egazette.nic.in/WriteReadData/2021/225464.pdf. Acesso em: 19 maio 2021.

INGRAM, David. **Facebook names 20 people to its 'Supreme Court' for content moderation**. NBC News. 6 maio 2020. Disponível em: https://www.nbcnews.com/tech/tech-news/facebook-names-20-people-its-supreme-court-content-moderation-n1201181. Acesso em: 07 maio 2021.

INSTAGRAM. **Central de Ajuda**. Disponível em: https://help.instagram.com/613868662393739 Acesso em : 08 fev. 2021.

INSTAGRAM. **Fornecer informação, segurança e apoio às pessoas no Instagram**. 24 mar. 2020. Disponível em: https://about.instagram.com/pt-br/blog/announcements/coronavirus-keeping-people-safe-informed-and-supported-on-instagram Acesso em: 18 jun. 2021.

IRIS - Instituto de Referência em Internet e Sociedade. **Transparência na moderação de conteúdo danoso:** lei, usuários e plataformas. Disponível em: https://irisbh.com.br/projetos/transparencia-na-moderacao-de-discurso-de-odio-lei-usuarios-e-plataformas/. Acesso em: 02 jan. 2020.

KAYE. David. **Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression**. 01 de junho de 2017. Disponível em: https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf. Acesso em: 17 jun. 2021

KELLER, Daphne. The Right Tools: Europe's Intermediary Liability Laws and the

EU 2016 General Data Protection Regulation. **Berkeley Technology Law Journal**, v. 33, p. 287, 2018. Disponível em: https://btlj.org/data/articles2018/vol33/33_1/Keller_Web.pdf Acesso em: 27 jun. 2021.

KENYALAW. Bloggers Association of Kenya (BAKE) v Attorney General & 3 others; Article 19 East Africa & another (Interested Parties) [2020] eKLR. Petition n. 206 of 2019. **Judgment**. Disponível em: http://kenyalaw.org/caselaw/cases/view/191276/ Acesso em: 15 jun. 2021. par. 45

KFCB - Kenya Film Classification Board. **Statement by the Kenya Film Classification Board Chief Executive Officer Dr. Ezekiel Mutua**, MBS on the banning of the film "Rafiki" issued on 27th april 2018. KFCB. 27 abr. 2018. Disponível em: http://kfcb.co.ke/wp-content/uploads/2018/04/CEO-STATEMENT-ON-RAFIKI-RESTRICTION-27-4-2018.pdf Acesso em: 7 maio 2021.

KIMUYU, Hillary. Blogger Alai charged for publishing 'alarming' claims on coronavirus. **Nairobi News**. 23 mar 2020. Disponível em: https://nairobinews.nation.co.ke/news/blogger-alai-charged-for-publishing-alarming-claims-on-coronavirus. Acesso em: 7 maio 2021.

KIMUYU, Hillary; KIPLAGAT, Sam. Court lifts ban on Rafili, film to run for an Oscar. **Nation.** 28 jun. 2020. Disponível em: https://nation.africa/kenya/news/court-lifts-ban-on-rafiki-film-to-run-for-an-oscar-89512 Acesso em: 07 maio 2021.

KLEINA, Nilton. Índia bane TikTok e mais 58 apps chineses por 'ameaça à soberania'. **Tecmundo**, 29 jun. 2020. Disponível em: https://www.tecmundo.com.br/software/154648-india-bane-tiktok-58-apps-chineses-ameaca-soberania.htm Acesso em 19 maio 2021.

KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech. **Harvard Law Review**, v. 131. n. 6. p. 1598-1670, abril de 2018. Disponível em: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Acesso em: 17 jun. 2021. p. 1625-1630.

KUMAR, Ankit; PATEL, Anand. Enforce policies regardless of political position: Under heat, Facebook denies ties with BJP in India. **India Today**. 16 ago. 2020. Disponível em: https://www.indiatoday.in/india/story/enforce-policies-regardless-political-position-under-heat-facebook-bjp-india-1711873-2020-08-16. Acesso em 19 maio 2021.

KURTZ, Lahis; VIEIRA, Victor. **Obtenção transnacional de conteúdo de comunicações telemáticas na América Latina:** relatório de pesquisa. Instituto de Referência em Internet e Sociedade: Belo Horizonte, 2020. Disponível em: https://irisbh.com.

br/wp-content/uploads/2020/04/Obten%C3%A7%C3%A3o-transnacional-de-conte%C3%BAdo-de-comunica%C3%A7%C3%B5es-telem%C3%A1ticas-na-Am%C3%A9rica-Latina.pdf Acesso em: 18 jun. 2021. p. 8

LAGEDO, Diego. 'Censura' a Donald Trump atenta contra a liberdade de expressão**.** **Blog de Jamildo**. 9 jan. 2021. Disponível em: https://blogs.ne10.uol.com.br/jamildo/2021/01/09/censura-a-donald-trump-atenta-contra-a-liberdade-de-expressao-por-diego-lagedo/ Acesso em: 11 maio 2021.

LISBOA, Alveni. Twitter constata que seu algoritmo de corte de fotos é mesmo racista. **Canaltech,** 20 maio 2021. Disponível em: https://canaltech.com.br/apps/twitter-constata-que-seu-algoritmo-de-corte-de-fotos-e-mesmo-racista-185388/ Acesso em: 09 jun 2021.

LOC - The Law Library of Congress. **Laws on Homosexuality in African Nations**. fev. 2014. Disponível em: https://www.loc.gov/law/help/criminal-laws-on-homosexuality/homosexuality-laws-in-african-nations.pdf Acesso em 14 jun. 2021.

MACKINNON, Rebecca; HICKOK, Elonnai; BAR, Allon; LIM, Hae-In. **Fostering freedom online:** the role of internet intermediaries. UNESCO series on internet freedom. França, 2014. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000231162_eng Acesso em: 14 jun. 2021.

MAOMBO, Sharon. DCI declares blogger Cyprian Nyakundi a wanted man. **Star**. 21 abr. 2020. Disponível em: https://www.the-star.co.ke/news/2020-04-21-dci-declares-blogger-cyprian-nyakundi-a-wanted-man/. Acesso em: 7 maio 2021.

MICHOUD, Adeline. Corporate Accountability in International Law: A'State Shaming'Exercise. **SOAS Law Journal,** v. 7, p. 190, 2020. Disponível em: https://8c607a4d-3252-415f-9696-9203fde0b41e.filesusr.com/ugd/79f2dc_ce5b501dbccc49f98cc6b3ae35d59830.pdf Acesso em: 18 jun. 2021.

MURAYA, David. Blogger Robert Alai Arrested For Posting Gory Photos. **Capital News**. 18 jun. 2019. Disponível em: https://www.capitalfm.co.ke/news/2019/06/blogger-robert-alai-arrested-for-posting-gory-photos/ Acesso em: 7 maio 2021.

MUTUA, Ezekiel. Google Takes Down Ethic's Raunchy Tarimbo Song Despite Apology. **Daily Active.** 9 set. 2019. Disponível em: https://www.dailyactive.info/2019/11/09/google-takes-down-ethics-raunchy-tarimbo-song-despite-apology/. Acesso em 19 maio 2021.

MUYA, Catherine. Policy Analisys on the Kenya Information and Communications Act. **Lawyers Hub.** 28 out. 2019. p. 18. Disponível em: https://lawyershub.org/media/

Policy_Analysis_on_Kenya_Information_and_Communcation_Amendment_Act_1_UXKIfMV.pdf. Acesso em: 10 maio 2021.

MYERS WEST, Sarah. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. **New Media & Society**, v. 20, n. 11, p. 4366-4383, 2018.

NCIC - National Cohesion and Integration Commission. **Guidelines on prevention of dissemination of undesirable bulk and premium rate political messages and political social media content via electronic communications networks** (NCIC Guidelines). jul. 2017. Disponível em: https://ca.go.ke/wp-content/uploads/2018/02/Guidelines-on-Prevention-of-Dissemination-of-Undesirable-Bulk-and-Premium-Rate-Political-Messages-and-Political-Social-Media-Content-Via-Electronic-Networks-1.pdf Acesso em 19 maio 2021.

NDAVULA, John; MUENI, Joy. New Media and Political Marketing in Kenya: The Case of 2013 General Elections. **International Journal of Arts and Commerce.** 3. 2014. p. 72. Disponível em: https://www.researchgate.net/publication/316189367_New_Media_and_Political_Marketing_in_Kenya_The_Case_of_2013_General_Elections Acesso em: 18 jun. 2021

NEWTON, Casey. **The Platformer.** Twitter Dumps Trump. Disponível em: https://www.platformer.news/p/-twitter-dumps-trump Acesso em: 03 fev. 2020.

NYABOLA, Nanjala. Putting Kenya's media shutdown in context. **Aljazeera**. 2 fev. 2018. Disponível em: https://www.aljazeera.com/opinions/2018/2/2/putting-kenyas-media-shutdown-in-context/ Acesso em 10 maio 2021.

OBIERA, Gladys Mokeira. Ethic Entertainment's Tarimbo ban - Why and fan reactions. [out. 2020]. **Tuko.** Disponível em: https://www.tuko.co.ke/322781-ethic-entertainments-tarimbo-ban-why-fan-reactions.html. Acesso em: 07 maio 2021.

OMONDI, Ian. Another suspect arrested for allegedly sending Robert Alai photos of dead officers. **Citizen Digital.** 19 jun. 2019. Disponível em: https://citizentv.co.ke/news/another-suspect-arrested-for-allegedly-sending-robert-alai-photos-of-dead-officers-259094/. Acesso em: 7 maio 2021.

O'NEILL, Onora. **A question of trust:** The BBC Reith Lectures 2002. Cambridge University Press, 2002.

ONIAS, Clarice Barbosa; RIBEIRO, Pedro Feliú. O impacto da presença chinesa na elite política do Quênia. **12º Encontro da ABCP.** 2020. Disponível em: https://cienciapolitica.org.br/system/files/documentos/eventos/2021/01/impacto-presenca-chinesa-elite-politica-quenia-3022.pdf Acesso em: 14 jun. 2021. p. 9

ONU - Organização das Nações Unidas. **Princípios Orientadores sobre Empresas e Direitos Humanos.** Implementando os Parâmetros "proteger, respeitar e reparar" das Nações Unidas. 16 jun. 2011. Trad. Ministério das Relações Exteriores. Disponível em: https://www.gov.br/mdh/pt-br/assuntos/noticias/2019/outubro/Cartilha_versoimpresso.pdf Acesso em 11 maio 2021.

ONU - Organização das Nações Unidas. **Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.** Report on Content Regulation. Symbol A/HRC/38/35. 6 abr. 2018. Disponível em: https://www.undocs.org/A/HRC/38/35 Acesso em: 14 jun. 2021. p. 16

PAES, Natiele. **Algoritmos:** o que são e como fazem parte do cotidiano? Politize. 7 maio 2020. Disponível em: https://www.politize.com.br/algoritmos-o-que-sao/ Acesso em 11 maio 2021.

PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones sobre gobernanza de contenido:** una guía para legisladores, reguladores y encargados de políticas empresariales. Access Now, 2020. Disponível em: https://www.accessnow.org/cms/assets/uploads/2020/03/Recomendaciones-Gov-Contenidos.pdf Acesso em: 18 jun. 2021.

PURNELL, Newley; HORWITZ, Jeff Facebook's Hate-Speech Rules Collide With Indian Politics. **The Wall Street Journal,** 14 ago. 2020. Disponível em: https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346. Acesso em 19 maio 2021.

QUÊNIA. Computer Misuse and Cybercrimes Act. **Kenya Gazette Suplement**. Nairobi, 16 maio 2018. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/ComputerMisuseandCybercrimesActNo5of2018.pdf . Acesso em 19 maio 2021.

QUÊNIA. Films and stage plays act. **National Council for Law Reporting.** 2012. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/FilmsandStagePlaysAct_Cap222.pdf. Acesso em 17 jun. 2021.

QUÊNIA. Information and Communications Act (Amendment Bill). 2019. **Kenya Gazette Suplement.** 24 jul. 2019, Nairobi. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/bills/2019/TheKenyaInformationandCommunication__Amendment_Bill_2019_NA_Bills_No._61.pdf Acesso em 14 jun. 2021. Parte I, seção 2.

QUÊNIA. Information and Communications Act (Amendment Bill). 2019. **Kenya Gazette Suplement**. 24 jul. 2019, Nairobi. Disponível em: http://kenyalaw.org/kl/fileadmin/pdfdownloads/bills/2019/TheKenyaInformationandCommunication__Amendment_Bill_2019_NA_Bills_No._61.pdf Acesso em 14 jun. 2021.

REISMAN, Dillon et al. Algorithmic impact assessments: A practical framework for public agency accountability. **AI Now Institute**, p. 1-22, 2018. Disponível em: https://ainowinstitute.org/aiareport2018.pdf. Acesso em 13 maio 2021.

ROBERTS, Sarah T. Digital detritus: 'Error' and the logic of opacity in social media content moderation. **First Monday.** 2018.

ROBERTS, Sarah T. **Behind the Screen:** Content Moderation in the Shadows of Social Media. New Haven: Yale University Press, 2019

RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 18 jun. 2021.

ROSHAN, Utkarsh. After India and Australia Controversies, Facebook Changes Terms and Conditions Again. **The Citizen,** 16 set. 2020. Disponível em: https://www.thecitizen.in/index.php/en/NewsDetail/index/9/19369/After-India-and-Australia-Controversies-Facebook-Changes-Terms-and-Conditions-Again. Acesso em 19 maio 2021.

RUTENBERG, Isaac; SUGOW, Abdulmalik. Regulation of the Social Media in Electoral Democracies: A Case of Kenya. **SOAS Law Journal**, v. 7, p. 301, 2020. Disponível em: https://www.cipit.strathmore.edu/wp-content/uploads/2020/06/Journal-7.pdf Acesso em: 18 jun. 2021. p. 321.

SANDER, Barrie. Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation. **Fordham International Law Journal**, v. 43, p. 939, 2019. Disponível em: https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2787&context=ilj. Acesso em: 11 maio 2021

SANDOVAL, Pablo Ximénez. Cresce o boicote de grandes anunciantes contra o Facebook apesar da reação de Zuckerberg. **El País**. 29 jun 2019. Disponível em: https://brasil.elpais.com/tecnologia/2020-06-29/cresce-o-boicote-de-grandes-anunciantes-contra-o-facebook-apesar-da-reacao-de-zuckerberg.html. Acesso em: 11 maio 2021.

SCHWENKE, Ken. Internet Company That Does Business With Hate Sites Alters Complaint Policies. **Propublica**. 10 maio 2017. Disponível em: https://www.propublica.org/article/internet-company-does-business-with-hate-sites-alters-complaint-policies. Acesso em: 18 jun. 2021.

SCHWENKE, Ken. Internet Company That Does Business With Hate Sites Alters Complaint Policies. **Propublica**. 10 maio 2017. Disponível em: https://www.propublica.org/article/internet-company-does-business-with-hate-sites-alters-complaint-policies. Acesso em: 18 jun. 2021.

SHALDERS, André. Como Brasil deixou de ser um dos campeões em liberdade na internet e se tornou 'parcialmente livre'. **BBC News**. 28 out. 2020. Disponível em: https://www.bbc.com/portuguese/brasil-54714957 Acesso em 19 maio 2021.

SILVA, Felipe Rocha et al. **NOTA TÉCNICA PL 2.630/2020 sobre a inclusão de mecanismos de transparência algorítmica no projeto de lei**. LAPIN, 2020. Disponível em: https://lapin.org.br/wp-content/uploads/2020/08/LAPIN-NT-PL-2630-Transparencia-Algoritmica.pdf Acesso em 11 maio 2021.

SNOWDEN, Edward. **Carta aberta ao povo do Brasil.** Disponível em: https://www1.folha.uol.com.br/mundo/2013/12/1386291-leia-integra-da-carta-de-snowden-ao-brasil.shtml Acesso em: 19 maio 2021.

SORRENTINO, Eduardo. Índia X Twitter: Rede social resiste às mudanças de privacidade. **Olhar Digital**. 6 jun. 2021. Disponível em: https://olhardigital.com.br/2021/06/06/internet-e-redes-sociais/india-x-twitter-rede-social-resiste-as-mudancas-de-privacidade/. Acesso em 14 jun. 2021.

SOUZA, Carlos Affonso de. Brasil não precisa importar nova regra de Trump sobre redes sociais. **Tilt**, 29 maio 2020. Disponível em: https://tecfront.blogosfera.uol.com.br/2020/05/29/brasil-nao-precisa-importar-nova-regra-de-trump-sobre-redes-sociais/?cmpid=copiaecola Acesso em: 11 jun. 2021.

SRINIVASAN, Kumar Bhargav et al. Content removal as a moderation strategy: Compliance and other outcomes in the changemyview community. **Proceedings of the ACM on Human-Computer Interaction**, v. 3, n. CSCW, p. 1-21, 2019. p. 2-3. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3359265. Acesso em: 03 fev. 2021.

STEIGER, Miriah et al. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support . **CHI '21:** Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, maio 2021. Disponível em: https://crowd.cs.vt.edu/wp-content/uploads/2021/02/CHI21_final__The_Psychological_Well_Being_of_Content_Moderators-2.pdf. Acesso em 14 jun. 2021.

STOP Hate for Profit. Disponível em: https://www.stophateforprofit.org/. Acesso em: 02 mar. 2021.

SWENEY, Mark. Mums Furious as Facebook Removes Breastfeeding Photos. **The Guardian**. 30 Dez. 2008. Disponível em: https://www.theguardian.com/media/2008/dec/30/facebookbreastfeeding-ban [https://perma.cc/2Y2B-9HDA. Acesso em: 02 mar. 2021.

TATSCH, Constança. Gordofobia: Instagram muda regras sobre exibição de seios após protesto. **Extra**. 26 out. 2020. Disponível em: https://extra.globo.com/noticias/gordofobia-instagram-muda-regras-sobre-exibicao-de-seios-apos-protesto-24713044.html Acesso em: 11 maio 2021.

TECH DESK. Who is Sudhir Krishnaswamy from Facebook's oversight board? **The Indian Express**. 8 maio 2020. Disponível em: https://indianexpress.com/article/technology/tech-news-technology/sudhir-krishnaswamy-facebook-oversight-board-6398217/ Acesso em: 18 jun. 2021.

TOSCANO, Eliphas Neto Palitot; DE MORAIS, Annaís Moraes. O Marco Civil da Internet como novo paradigma para a responsabilidade civil dos provedores de serviços da rede: o judicial notice and takedown. In: **Direito e Novas Tecnologias II**. SIMÃO FILHO, Adalberto; PEREIRA JÚNIOR, Antonio Jorge, NASCIMENTO, Valéria Ribas do (org.). – Florianópolis : CONPEDI, 2014. Disponível em: http://www.publicadireito.com.br/artigos/?cod=93c442d40a4e0b6f Acesso em: 18 jun. 2021.

TRIMBLE, Marketa. Geoblocking, Technical Standards and the Law. In: LOBATO, Ramon; MEESE, James (org.) **Geoblocking and global video culture**. 2016. Disponível em: http://scholars.law.unlv.edu/facpub/947. Acesso em: 09 fev 2021.

TWITTER. **Central de ajuda.** Disponível em: https://help.twitter.com/pt/rules-and-policies/twitter-rules Acesso em: 11 maio 2021.

TWITTER. Permanent suspension of @realDonaldTrump. **Twitter Blog**, 08 de janeiro de 2021. Disponível em: https://blog.twitter.com/en_us/topics/company/2020/suspension. Acesso em: 17 jun. 2021.

TWOREK, Heidi; LEERSSEN, Paddy. **An Analysis of Germany's NetzDG Law**. Abr. 2019. p. 4-8. Disponível em: https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf. Acesso em: 19 maio 2021.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). **Jornal Oficial da União Europeia**. 4 maio 2016. Disponível em: https://gdpr-info.eu/ Acesso em: 11 maio 2021.

WAKEFIELD, Jane. 'Após trabalhar como moderadora na web, parei de apertar a mão das pessoas – fiquei com nojo da humanidade'. **BBC News**. 15 out. 2018. Disponível em: https://www.bbc.com/portuguese/geral-45861208. Acesso em 14 jun. 2021.

WONG, Julia. How Facebook let fake engagement distort global politics: a whistleblower's account. **The Guardian**, 12 abr. 2021. Disponível em: https://www.theguardian.com/technology/2021/apr/12/facebook-fake-engagement-whistleblower-sophie-zhang. Acesso em 16 jun. 2021.

YALÇIN, Nursel; KÖSE, Utku. What is search engine optimization: SEO?. **Procedia-Social and Behavioral Sciences**, v. 9, p. 487-493, 2010. p. 488-489. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877042810022901. Acesso em: 03 fev. 2021.

YOUTUBE. **Diretrizes da comunidade**. Disponível em: https://www.youtube.com/howyoutubeworks/policies/community-guidelines/#community-guidelines. Acesso em: 03 fev. 2021.

ZUBOFF, Shoshana. Big other: surveillance capitalism and the prospects of an information civilization. **Journal of information technology**, n. 30. Londres, 2015. Disponível em: https://cryptome.org/2015/07/big-other.pdf. Acesso em: 13 maio 2021

APÊNDICE A - Requisitos de transparência na moderação de conteúdo

Esta é uma síntese dos requisitos a serem observados para elaborar regulação dirigida à atividade de moderação de conteúdo, de acordo com os documentos e padrões internacionais levantados nesta pesquisa. Embora não seja exaustiva, considera-se que essa lista compreende as principais medidas de transparência recomendadas. Os documentos selecionados contém ainda outros requisitos, relativos a outros aspectos da atividade; para esta pesquisa, apenas os que diziam respeito à transparência foram elencados.

Os requisitos mínimos de **restrição de conteúdos considerados ilegais** seriam:

• A base legal para a afirmação de que o conteúdo é ilegal; • O identificador de Internet e descrição do conteúdo alegadamente ilegal; • Considerações relativas a limitações, exceções e defesas disponíveis para o usuário provedor do conteúdo que pode ser afetado; • Informações de contato da parte emissora da solicitação ou de seus representantes, a menos que isso seja proibido por lei; • Evidências suficientes para documentar as bases legais para a apresentação do pedido; • Uma declaração de boa-fé indicando que as informações fornecidas estão corretas.	Princípios de Manila, 3.b[1]
• Se a base for uma ordem de autoridade (judicial, por exemplo), a restrição deve constar em relatório	Princípios de Manila, 3.f[2]
• A plataforma deve apresentar informação ao criador do conteúdo sobre qual medida será tomada	Access Now, p. 41[3]

1 EFF - Electronic Frontier Foundation et al. **Princípios de Manila sobre responsabilidade civil de intermediários**. Disponível em: https://manilaprinciples.org/pt-br/principles.html Acesso em: 14 jun. 2021.

2 EFF - Electronic Frontier Foundation et al. **Princípios de Manila [...]** op. cit.

3 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones sobre gobernanza de contenido**:

Os requisitos mínimos de **restrição com base em políticas de comunidade** seriam:

• As razões pelas quais o conteúdo em questão viola as políticas de restrição de conteúdos do intermediário; • O identificador de Internet e descrição da alegada violação das políticas de restrição de conteúdo; • Informações de contato da parte emissora da solicitação ou de seus representantes, a menos que isso seja proibido por lei; • Uma declaração de boa-fé indicando que as informações fornecidas estão corretas.	Princípios de Manila, 3.c[4]
• Link, trecho ou outra informação do conteúdo que permita identificar o que será removido/restringido • Cláusula específica das políticas que foram consideradas violadas pelo conteúdo. • Como o conteúdo foi detectado ou restringido (se denunciado por usuário, pelo governo [quando autorizado por lei] ou por terceiros confiáveis, detecção automatizada, ou reclamação externa)	Princípios de Santa Clara, 2[5]
• Informação sobre qual medida será tomada	Access Now, p. 41[6]

Em **todos os casos de restrição**, recomenda-se informar ao usuário afetado sobre seus direitos, incluindo mecanismos de contra-notificação ou recurso (Princípios de Manilla, 3.e; Princípios de Santa Clara, 2), bem como notificação dos resultados da revisão com exposição suficiente de motivos para a decisão (Princípios de Santa Clara, 3).

Nos **relatórios de transparência**, recomenda-se que deve constar:

una guía para legisladores, reguladores y encargados de políticas empresariales. Access Now, 2020. Disponível em: https://www.accessnow.org/cms/assets/uploads/2020/03/Recomendaciones-Gov-Contenidos.pdf. Acesso em: 18 jun. 2021.

4 EFF - Electronic Frontier Foundation et al. **Princípios de Manila [...]** op. cit.

5 EFF - Electronic Frontier Foundation et al. **Santa Clara Principles on transparency and accountability in content moderation**. Disponível em: https://santaclaraprinciples.org/. Acesso em: 14 jun. 2021.

6 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones [...]** op. cit.

• Informações específicas sobre todas as restrições de conteúdos realizadas pelo intermediário, incluindo ações tomadas devido à requisição governamental, ordens judiciais, requisições de agentes privados e a implementação de políticas de restrição de conteúdo;	Princípios de Manila, 6.e[7]
• Número total de postagens e contas denunciadas e suspensas ou removidas, por tipo de regra violada, por formato ou conteúdo (ex: texto, áudio, vídeo...), por fonte da denúncia (ex: governos, terceiros confiáveis, detecção automatizada, etc.), por localização dos denunciantes e dos usuários impactados [onde visível].	Princípios de Santa Clara, 1[8]
• Prazos usuais para envio de notificação ao provedor de conteúdo antes de adotar uma medida, tempo transcorrido antes da restrição ao conteúdo, prazo para uma apelação e quantidade de recursos recebidos sobre medidas ou notificações	Access Now, p. 36[9]

Também é recomendada a **publicação** de:

• Políticas de restrição de conteúdos online com uma linguagem clara e em formatos acessíveis, atualizadas na medida em que evoluem, notificando os usuários das alterações quando aplicável; • Uso, nelas, de termos explícitos e detalhados, especialmente em relação ao que é considerado violação de política, as sanções correspondentes e meios de recurso	Princípios de Manila, 6.c[10] ONU Discurso de Ódio, p. 22[11] Guia do Conselho da Europa[12] Princípios Ruggie, p. 36[13] Access Now, p. 39[14]

7 EFF - Electronic Frontier Foundation et al. **Princípios de Manila [...]** op. cit.

8 EFF - Electronic Frontier Foundation et al. **Santa Clara Principles [...]** op. cit.

9 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones [...]** op. cit.

10 EFF - Electronic Frontier Foundation et al. **Princípios de Manila [...]** op. cit.

11 ONU - Organização das Nações Unidas. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. **Report on Hate Speech**. Symbol A/74/486. 9 out. 2019. Disponível em: https://undocs.org/A/74/486. Acesso em: 17 jun. 2021.

12 CE - Conselho da Europa. **Guia dos Direitos Humanos para os Utilizadores de Internet.** 16 abr. 2014. Disponível em: http://www.odionao.com.pt/media/5185/GuiaDireitosHumanosUtilizadoresInternet.pdf Acesso em 11 maio 2021. p. 57

13 ONU - Organização das Nações Unidas. **Princípios Orientadores sobre Empresas e Direitos Humanos**. Implementando os Parâmetros "proteger, respeitar e reparar" das Nações Unidas. 16 jun. 2011. Trad. Ministério das Relações Exteriores. Disponível em: https://www.gov.br/mdh/pt-br/assuntos/noticias/2019/outubro/Cartilha_versoimpresso.pdf Acesso em 11 maio 2021.

14 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones [...]** op. cit.

• Uma notificação explícita que explique ao usuário que tenta acessar o conteúdo suspenso/removido qual conteúdo foi restrito e o motivo para tanto quando possível	Princípios de Manila, 6.c[15]
• Relatórios detalhados, pelo Estado, de transparência sobre pedidos relativos a conteúdo feitos a intermediários. • Mecanismos de criação, funcionamento e avaliação de ferramentas automatizadas de moderação de conteúdo	ONU Regulação, par. 69[16]
• Informação sobre leis fora dos Estados Unidos da América que tenham impacto na liberdade de expressão	Access Now, p. 33[17]
• Explicações sobre como os pedidos governamentais em relação a conteúdo são encaminhados	GNI, p. 98[18]
• Informação sobre as formas de treinamento e condições de contratação e de trabalho da equipe de moderação de conteúdo, ou sobre as formas de testagem e treinamento de algoritmos de moderação automatizada.	IRIS[19]

Outra recomendação geral de transparência é que haja **abertura para participação** do público no cenário regulatório sobre moderação de conteúdo[20 21 22].

Essas medidas possibilitam o compartilhamento, pelo setor privado, de motivos, procedimentos e critérios com o público. Por isso, são importantes ferramentas para a elaboração multissetorial de práticas de moderação adequadas a direitos humanos.

15 EFF - Electronic Frontier Foundation et al. **Princípios de Manila [...]** op. cit.

16 ONU - Organização das Nações Unidas. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. **Report on Content Regulation**. Symbol A/HRC/38/35. 6 abr. 2018. Disponível em: https://www.undocs.org/A/HRC/38/35. Acesso em: 17 jun. 2021.

17 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones [...]** op. cit.

18 GNI - Global Network Initiative. **The GNI Principles Principles on Freedom of Expression and Privacy**. Disponível em: https://globalnetworkinitiative.org/gni-principles/ Acesso em: 14 jun. 2021.

19 RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: https://bit.ly/3nUbXYh. Acesso em: 19 nov. 2020. p. 68, 76.

20 ONU - Organização das Nações Unidas. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. **Report on Content Regulation**. op. cit. par. 69

21 GNI - Global Network Initiative. **The GNI Principles Principles [...]** op. cit. p. 98

22 PIRKOVÁ, Eliška; PALLERO, Javier. **26 recomendaciones [...]** op. cit. p. 44

APÊNDICE B - Quadro de dispositivos regulatórios analisados

PAÍS	REGULAÇÃO - NOME ORIGINAL	REGULAÇÃO - NOME TRADUZIDO ADOTADO NO ESTUDO
ALEMANHA	Telekommunikationsgesetz	Lei de Telecomunicações
	Urheberrechtsgesetz	Lei de Direitos Autorais e Direitos Correlatos
	Netzdurchsetzunggesetz	Lei de Regulação da Internet
	Strafgesetzbuch	Código Criminal
	Medienstaatsvertrag	Tratado de Mídia Estatal
	Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität	Lei para Combate ao Extremismo de Direita e ao Crime de Ódio
AUSTRÁLIA	Telecommunications and Other Legislation Amendment (Assistance and Access) Act	Lei para Emenda de Legislação Relativa a Telecomunicações e Outros (Assistência e Acesso)
	Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act	Emenda ao Código Penal (Compartilhamento de Material Abominavelmente Violento)
	Treasury Laws Amendment (News Media and Digital Platforms Mandatory Bargaining Code) Bill	Lei de Emenda às Leis do Tesouro (Código de Negociação Obrigatório para Plataformas Digitais e Mídia de Notícias)
	Online Safety Bill	Projeto de Lei para Segurança Online
	Copyright Act	Lei de Direitos Autorais

PAÍS	REGULAÇÃO - NOME ORIGINAL	REGULAÇÃO - NOME TRADUZIDO ADOTADO NO ESTUDO
BRASIL	Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil	Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil
	Projeto de Lei nº 2.630, 03 jul. 2020. (Do Sr. Alessandro Vieira). Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet	Projeto de Lei nº 2.630, 03 jul. 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet
	Projeto de Lei nº 21, 04 fev. 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências	Projeto de Lei nº 21, 04 fev. 2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências
ÍNDIA	Information Technology – Guidelines For Intermediaries And Digital Media Ethics Code – Rules, 2021	Regras de Tecnologia da Informação – Diretrizes do Intermediário e Código de Ética da Mídia Digital 2021
QUÊNIA	Computer Misuse and Cibercrimes Act	Lei de Cibercrimes e Uso Indevido de Computadores
	Guidelines on prevention of dissemination of undesirable bulk and premium rate political messages and political social media content via electronic communications networks (NCIC Guidelines)	Diretrizes da NCIC
	Films and stage plays act	Lei de Filmes e Peças de Teatro
	Information and Communications Act (Amendment Bill)	Projeto de Emenda da Lei de Informações e Comunicações

APÊNDICE C - Formulário de análise regulatória

Formulário análise regulatória

Formulário com as perguntas e campos a serem preenchidos para analisar:

- se há regulação sobre moderação de conteúdo na jurisdição analisada
- os focos de atenção da regulação existente sobre o tema
- a relevância dada à transparência sobre moderação de conteúdo pelos instrumentos regulatórios
- as lacunas da regulação quanto à transparência na moderação de conteúdo

*Obrigatório

País *

Sua resposta

1) Quais preocupações o país apresenta sobre conteúdo online?

- [ ] Atenção midiática
- [ ] Debates no legislativo
- [ ] Processo judicial em andamento
- [ ] Jurisprudência
- [ ] Diretrizes não-legislativas de órgãos não-oficiais
- [ ] Diretrizes não-legislativas de órgãos oficiais
- [ ] Projeto de lei
- [ ] Lei
- [ ] Outro: _____

2) Que documentos ou instrumentos regulatórios foram identificados?

Formato da resposta: Tipo de instrumento (Notícia/Diretriz/Projeto/Lei) - Nome do instrumento - link

Sua resposta

3) O país demonstra preocupação regulatória (projetos/leis/normas ou diretrizes) sobre conteúdo online? *

Caso só haja jurisprudência, sem diretrizes, projetos, leis e afins, assinalar "não".

- ◯ Sim
- ◯ Não

Próxima

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

Google Formulários

Formulário análise regulatória

Cenário regulatório

4) O cenário regulatório sobre moderação de conteúdo conta com (abertura para) participação pública?

- ◯ Sim
- ◯ Não

5) Quais as evidências de abertura ou participação pública?

Nome da fonte (Título da notícia e jornal/site - link; título do documento e autor - link)

Sua resposta

6) A lei/diretriz/projeto

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] restringe a possibilidade de decidir sobre a legalidade de um discurso às autoridades judiciais
- [ ] estabelece que a plataforma ou terceiros devem monitorar ou filtrar proativamente conteúdo
- [ ] prevê que o Estado deve publicar relatórios detalhados de transparência sobre pedidos relativos a conteúdo feitos a intermediários
- [ ] define os limites à liberdade de expressão, com conceitos-chave
- [ ] impõe que a plataforma remova conteúdo dentro de determinado prazo
- [ ] atribui a terceiros a responsabilidade por conteúdo gerado por usuário
- [ ] Nehuma das alternativas
- [ ] Outro: _____

7) A lei/diretriz/projeto obriga as plataformas a

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] publicizar os meios e critérios de detecção e identificação de conteúdo irregular
- [ ] usarem termos explícitos e detalhados sobre o tipo de conteúdo considerado danoso ou sujeito a restrição
- [ ] notificar o usuário a respeito de intervenção sobre conteúdo que este tenha gerado/compartilhado
- [ ] informar o usuário sobre o motivo de medidas interventivas sofridas em seu conteúdo
- [ ] oferecer ao usuários mecanismo de questionamento de medida interventiva adotada sobre seu conteúdo
- [ ] disponibilizar ferramentas para apontar falhas em mecanismos automatizados de moderação de conteúdo
- [ ] divulgar relatórios de transparência sobre moderação de conteúdo realizada com base em suas políticas
- [ ] Nenhuma das alternativas
- [ ] Outro: _____

8) Se a regulação analisada trata de restrição de conteúdo por ser considerado ilegal, ela determina como requisito para a medida:

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] Não trata do tópico
- [ ] A base legal para a afirmação de que o conteúdo é ilegal;
- [ ] O identificador de Internet e descrição do conteúdo alegadamente ilegal;
- [ ] Considerações relativas a limitações, exceções e defesas disponíveis para o usuário provedor do conteúdo que pode ser afetado;
- [ ] Informações de contato da parte emissora da solicitação ou de seus representantes, a menos que isso seja proibido por lei;
- [ ] Evidências suficientes para documentar as bases legais para a apresentação do pedido;
- [ ] Uma declaração de boa-fé indicando que as informações fornecidas estão corretas.
- [ ] Se a base for uma ordem de autoridade (judicial, por exemplo), a restrição deve constar em relatório (Princípios de Manilla, 3.f)
- [ ] A plataforma deve apresentar informação, ao criador do conteúdo, sobre qual medida será tomada (Access Now, p. 41)
- [ ] Outro: _____

9) Se a regulação analisada trata de restrição de conteúdo com base em política de comunidade, ela determina como requisito para a medida:

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] Não trata do tópico
- [ ] As razões pelas quais o conteúdo em questão viola as políticas de restrição de conteúdos do intermediário;
- [ ] O identificador de Internet e descrição da alegada violação das políticas de restrição de conteúdo;
- [ ] Informações de contato da parte emissora da solicitação ou de seus representantes, a menos que isso seja proibido por lei;
- [ ] Uma declaração de boa-fé indicando que as informações fornecidas estão corretas.
- [ ] Link, trecho ou outra informação do conteúdo que permita identificar o que será removido/restringido

- [ ] Cláusula específica das políticas que foram consideradas violadas pelo conteúdo.
- [ ] Como o conteúdo foi detectado ou restringido (se denunciado por usuário, pelo governo [quando autorizado por lei] ou por terceiros confiáveis, detecção automatizada, ou reclamação externa)
- [ ] Informação sobre qual medida será tomada (Access Now, p. 41)
- [ ] Outro: _____

10) Quanto a restrição de conteúdo em geral, a lei estabelece que a plataforma deve:

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] informar ao usuário afetado sobre seus direitos
- [ ] informar sobre mecanismos de contra-notificação ou recurso
- [ ] notificar dos resultados da revisão
- [ ] expor, na decisão do recurso, motivos suficientes para a decisão
- [ ] Nenhuma das alternativas
- [ ] Outro: _____

11) A lei determina que conste, nos relatórios de transparência:

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] Informações específicas sobre todas as restrições de conteúdos realizadas pelo intermediário, incluindo ações tomadas devido à requisição governamental, ordens judiciais, requisições de agentes privados e a implementação de políticas de restrição de conteúdo; (Princípios de Manilla, 6.e)
- [ ] Número total de postagens e contas denunciadas e suspensas ou removidas, por tipo de regra violada, por formato ou conteúdo (ex: texto, áudio, vídeo...), por fonte da denúncia (ex: governos, terceiros confiáveis, detecção automatizada, etc.), por localização dos denunciantes e dos usuários impactados [onde visível]. (Princípios de Santa Clara, 1)
- [ ] Prazos usuais para envio de notificação ao provedor de conteúdo antes de adotar uma medida, tempo transcorrido antes da restrição ao conteúdo, prazo para uma apelação e quantidade de recursos recebidos sobre medidas ou notificações (Access now, p. 36)

- [ ] Nenhuma das alternativas
- [ ] Outro: _____

12) A lei determina que as políticas de comunidade

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] Explicitem as políticas de restrição de conteúdos online
- [ ] Contenham explicitamente o que é considerado violação de política, as sanções correspondentes e meios de recurso
- [ ] Adotem linguagem clara, termos explícitos e detalhados
- [ ] Estejam formatos acessíveis
- [ ] Sejam mantidas atualizadas na medida em que evoluem, notificando-se os usuários das alterações quando aplicável
- [ ] Nenhuma das anteriores
- [ ] Outro: _____

13) A lei determina que haja a publicação de:

Utilizar o campo "outros" para anotar a lei e o dispositivo (artigo, parágrafo) onde está aquela informação.

- [ ] Uma notificação que explique [ao usuário que tenta acessar o conteúdo suspenso/removido] qual conteúdo foi restrito e o motivo para tanto [quando possível] (Princípios de Manilla, 6.f)
- [ ] Relatórios detalhados, pelo Estado, de transparência sobre pedidos relativos a conteúdo feitos a intermediários. (ONU, Relatoria sobre Regulação, par. 69)
- [ ] Mecanismos de criação, funcionamento e avaliação de ferramentas automatizadas de moderação de conteúdo (Access Now, p. 33)
- [ ] Informação sobre leis fora dos Estados Unidos da América que tenham impacto na liberdade de expressão
- [ ] Explicações sobre como os pedidos governamentais em relação a conteúdo são encaminhados (GNI, p. 98)
- [ ] Nenhuma das anteriores
- [ ] Outro: _____

Voltar Próxima

Formulário análise regulatória

*Obrigatório

Jurisprudência

Decisões com impacto sobre o cenário regulatório

14) Foi localizado algum caso judicial relevante a esse cenário regulatório? *

- ◯ Sim
- ◯ Não
- ◯ Outro: _____

15) Órgão julgador

Se houver mais de um, inserir todos, em ordem e numerados no formato 1); 2); um em cada linha (ctrl+enter para separar linhas)

Sua resposta

16) Resumo do(s) caso(s) e relevância

Responder enumerando se houver mais de uma; no mesmo formato e ordem da pergunta anterior. Informar data do início e término da ação ou do julgamento/decisão (indicar se ainda está em trâmite se for o caso), síntese da demanda, conexão com o tema, link

Sua resposta

Voltar | Próxima

Nunca envie senhas pelo Formulários Google.

Este conteúdo não foi criado nem aprovado pelo Google. Denunciar abuso - Termos de Serviço - Política de Privacidade

iris

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE