



Governança da moderação de conteúdo online

Percepções sobre o papel
dos atores e regimes

iris

INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE

Governança da moderação de conteúdo online

Percepções sobre o papel
dos atores e regimes

AUTORIA

Fernanda dos Santos Rodrigues Silva
Júlia Maria Caldeira Gertrudes

REVISÃO

Gustavo Ramos Rodrigues

REVISÃO EXTERNA

Pedro de Perdigão Lana

PROJETO GRÁFICO, CAPA, DIAGRAMAÇÃO E FINALIZAÇÃO

Felipe Duarte

PRODUÇÃO EDITORIAL

IRIS - Instituto de Referência em Internet e Sociedade

COMO CITAR EM ABNT

SILVA, Fernanda dos Santos Rodrigues; GERTRUDES, Júlia Maria Caldeira. **Governança da moderação de conteúdo online**: percepções sobre o papel dos atores e regimes. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2023. Disponível em: <<https://bit.ly/3WHMUlg>>. Acesso em: dd mmm aaaa.

ISBN: 978-65-87416-03-8



**INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE**

DIREÇÃO

Gustavo Rodrigues

Paloma Rocillo

MEMBROS

Ana Bárbara Gomes | Coordenadora de Políticas Públicas e Pesquisadora

Felipe Duarte | Coordenador de Comunicação

Fernanda Rodrigues | Coordenadora de Pesquisa e Pesquisadora

Juliana Roman | Pesquisadora

Júlia Caldeira | Pesquisadora

Lucas Samuel | Estagiário de pesquisa

Luiza Dutra | Pesquisadora

Paulo Rená da Silva Santarém | Pesquisador

Rafaela Ferreira | Estagiária de pesquisa

Thais Moreira | Estagiária de comunicação

Victor Barbieri Rodrigues Vieira | Pesquisador

Wilson Guilherme | Pesquisadore

irisbh.com.br

SUMÁRIO

1. INTRODUÇÃO	<u>6</u>
2. A GESTÃO DA MODERAÇÃO DE CONTEÚDO: ENTRE ATORES E REGIME	<u>8</u>
2.1. Atores envolvidos na gestão da moderação de conteúdo e os diferentes regimes	<u>8</u>
2.2. Modelo regulatório de moderação de conteúdo no Brasil	<u>13</u>
3. METODOLOGIA	<u>15</u>
3.1. Definição do caráter da entrevista e elaboração do roteiro de perguntas	<u>15</u>
3.2. Composição da amostra de pessoas entrevistadas	<u>17</u>
3.3. Procedimentos para a realização das entrevistas	<u>18</u>
3.4. Categorias e códigos para análise de conteúdo das entrevistas	<u>19</u>
3.5. Limitações da pesquisa	<u>19</u>
4. RESULTADOS: ANÁLISE DAS ENTREVISTAS	<u>20</u>
4.1. Sobre moderação de conteúdo e suas técnicas	<u>20</u>
Sobre as principais funções e ações da moderação de conteúdo	<u>20</u>
Função: Sanitização do espaço virtual	<u>21</u>
Função: Produto das plataformas	<u>22</u>
Ação: Intervenção	<u>24</u>
Ação: Curadoria de conteúdo	<u>25</u>
Sobre o uso de inteligência artificial para moderação de conteúdo	<u>29</u>
Sobre a necessidade de revisão humana de decisões automatizadas	<u>33</u>
4.2. Sobre o papel e atuação dos diferentes atores da governança da internet envolvidos na moderação de conteúdo	<u>35</u>
Sobre a responsabilidade pela gestão da moderação de conteúdo	<u>35</u>
Sobre possíveis meios de fiscalização do processo de moderação de conteúdo	<u>39</u>

Sobre instâncias privadas de revisão de moderação de conteúdo	<u>43</u>
Sobre o papel do Poder Judiciário enquanto possível instância de revisão judicial de decisões envolvendo moderação de conteúdo	<u>48</u>
Sobre a participação do terceiro setor e academia nos processos regulatórios de conteúdo online	<u>51</u>
4.3. Sobre a avaliação dos diferentes modelos regulatórios	<u>57</u>
Sobre a autorregulação por parte das plataformas na moderação de conteúdo	<u>57</u>
Sobre a regulação estatal na moderação de conteúdo	<u>60</u>
Sobre as recomendações sem efeito legal (Soft Law)	<u>63</u>
Sobre normas estatais com diretrizes procedimentais	<u>66</u>
Sobre normas estatais prevendo a remoção de conteúdos específicos	<u>68</u>
Sobre o regime de autorregulação regulada	<u>71</u>
4.4. Sobre a regulação da moderação de conteúdo no Brasil	<u>76</u>
Sobre o modelo regulatório de moderação de conteúdo vigente no Brasil	<u>76</u>
Sobre os aspectos positivos e negativos, assim como os impactos esperados a partir da aprovação do Projeto de Lei 2.630 (“PL das Fake News”)	<u>80</u>
5. DISCUSSÃO E CONSIDERAÇÕES FINAIS	<u>85</u>
6. REFERÊNCIAS BIBLIOGRÁFICAS	<u>94</u>
APÊNDICE I - ROTEIRO DA ENTREVISTA	<u>98</u>
Bloco I - Apresentação e temas gerais	<u>98</u>
Bloco II - Moderação de conteúdo e a interação entre os diferentes atores envolvidos	<u>98</u>
Bloco III - Modelos regulatórios	<u>99</u>
APÊNDICE II - LISTAGEM DE CÓDIGOS UTILIZADOS	<u>101</u>

1. Introdução

Quando o assunto é moderação de conteúdo, a atuação por parte das plataformas costuma ser o primeiro pensamento que vem à mente. No papel de intermediárias, são elas que estabelecem e aplicam as primeiras normas, definidas em termos de uso e diretrizes de comunidade, a fim de controlar seu ambiente online e garantir a liberdade de expressão dos usuários, ao mesmo tempo em que buscam protegê-los da exposição a conteúdos nocivos. Nesse sentido, devido ao poder de que dispõem, é possível afirmar que essas empresas de tecnologia reúnem em si uma dinâmica de governança centralizada que costuma ser dividida em três poderes em um Estado de direito: legislativo, executivo e judiciário.

Enquanto estabelecem as normas para que os usuários possam se comunicar, interagir e permanecer naquela comunidade, as plataformas também aplicam essas regras e tomam decisões com base nelas. Em um contexto de autorregulação, são os intermediários que se tornam o foco da governança de conteúdo online. Entretanto, olhando para o ecossistema da própria governança da internet, é possível observar que outros atores igualmente relevantes são capazes de influir no poder conferido às plataformas.

Normas estatais, por exemplo, somam-se não apenas a partir de uma regulação específica para o setor, mas também ao trazer definições de atividades ilegais que devem ser perseguidas e responsabilizadas em qualquer contexto. Sociedade civil e comunidade técnico-científica apresentam aportes críticos e contribuem para a evolução do debate público a respeito do tema, participando da construção de recomendações internacionais sobre moderação, como os Princípios de Manilla e Santa Clara. Mesmo os usuários têm um papel importante ao trazer feedbacks a partir do uso de determinada plataforma.

Mais recentemente, atores como instâncias privadas e independentes de revisão de decisões de moderação, a exemplo do Oversight Board,¹ da Meta, também têm se somado a esse cenário ao dividir a responsabilidade de tornar essa atividade mais transparente e alinhada a parâmetros internacionais de direitos humanos. Em verdade, a descentralização da tomada de decisão por parte das plataformas pode auxiliar para que estas não sejam as únicas a serem acionadas quando algum erro ou equívoco é constatado no processo de intervenção em conteúdo de terceiros.

1 O Oversight Board é um “comitê de supervisão independente, criado especialmente para receber apelações de usuários insatisfeitos com as decisões da empresa. (...) Segundo seu estatuto, o Oversight Board pode revisar apenas casos envolvendo conteúdos que tenham sido removidos por violar as políticas do Facebook”. In: ARHEGAS, João Victor; GODOY, Miguel Gualano de. Os limites da jurisdição do Facebook Oversight Board: de Marbury v Madison para Facebook v Trump. **Jota Info**, 02 de fevereiro de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/os-limites-da-jurisdicao-do-facebook-oversight-board-02022021>. Acesso em: 20 dez. 2022.

No Brasil, a discussão em torno da moderação ganhou ainda mais relevo desde as eleições de 2018, em que a corrida presidencial contou com a influência da disseminação massiva de notícias falsas,² tendo as redes sociais e aplicativos de mensageria privada como os principais veículos de propagação. Essa tendência, que já havia sido observada no contexto eleitoral estadunidense de 2016, com a eleição de Donald Trump,³ ressaltou a necessidade de se olhar para as plataformas com maior atenção e buscar entender como enfrentar a proliferação de conteúdo desinformativo.

Em vista disso, é necessário compreender melhor os aspectos da interação entre os diferentes atores envolvidos na regulação de conteúdo online, sendo essa compreensão essencial para pensar, inclusive, nas possibilidades de modelos regulatórios. Da criação de normas estatais à elaboração de recomendações sem efeito legal, existem outros mecanismos, que não aqueles derivados das plataformas, que podem impactar nos procedimentos de moderação e que, portanto, demandam atenção e um estudo mais aprofundado.

Nesse contexto, a presente pesquisa buscou identificar as crenças, percepções e argumentos em torno do papel dos diferentes setores na regulação da moderação de conteúdo (como é a participação de governos, empresas e sociedade civil na elaboração de regras e padrões) e sobre os entraves e vantagens de cada tipo de modelo regulatório na adequação da moderação a direitos humanos. Para isso, realizaram-se entrevistas com especialistas ou profissionais engajados com o tema de regulação de conteúdo online, advindos de diferentes setores.

Assim, a fim de apresentar os resultados encontrados, o presente trabalho contém, além desta introdução, mais quatro capítulos. O primeiro deles destina-se a apresentar o marco teórico voltado a situar o debate existente em torno dos regimes de moderação de conteúdo hoje em dia, em especial, no cenário brasileiro. O segundo é voltado para a descrição da metodologia empregada nesta pesquisa, assim como suas eventuais limitações. O terceiro capítulo é dedicado à análise de conteúdo das entrevistas e o último capítulo, por fim, destina-se a apresentar as considerações finais das autoras.

2 GAMA, Sophia. Guerra de desinformação: as fake news nas eleições de 2018. **Câmara Municipal de Curitiba**, 15 de julho 2022. Disponível em: <https://www.curitiba.pr.leg.br/informacao/noticias/guerra-de-desinformacao-as-fake-news-nas-eleicoes-de-2018>. Acesso em: 19 dez. 2022.

3 MARS, Amanda. Como a desinformação influenciou nas eleições presidenciais? **El País**, 25 de fevereiro 2018. Disponível em: https://brasil.elpais.com/brasil/2018/02/24/internacional/1519484655_450950.html. Acesso em: 19 dez. 2022.

2. A gestão da moderação de conteúdo: entre atores e regime

Conforme já mencionado, a gestão da moderação de conteúdo não se restringe ao papel desempenhado pelas plataformas online. A responsabilidade pelo controle do que é postado na rede divide-se também com outros atores da governança da internet, como é o caso do setor governamental, da sociedade civil e da comunidade técnico-científica, que também podem influenciar no processo regulatório de moderação.

Assim, o ecossistema de regulamentação pode ser originado a partir de diferentes fontes: a criação de normas estatais ou internacionais emitidas por autoridades públicas, sejam elas diretamente sobre a atividade das plataformas ou sobre a forma como elas devem proceder ao regular conteúdo; padrões e diretrizes com caráter de recomendação, emitidas por grupos de agentes não governamentais; além das próprias políticas de comunidade e termos de uso de plataformas sociais.

O maior ou menor grau de controle exercido tanto pelos diferentes agentes como pelas distintas fontes normativas impactará, por sua vez, na identificação do modelo de regime de moderação vigente em determinado contexto, podendo ser este autorregulatório, corregulatório ou heterorregulatório. Pensando nisso, a presente seção abordará brevemente o papel dos diferentes atores envolvidos na gestão de moderação de conteúdo, assim como das regras que compõem esse ambiente normativo.

2.1. Atores envolvidos na gestão da moderação de conteúdo e os diferentes regimes

Diferentes autores buscaram conceitualizar a governança de plataformas e os agentes envolvidos nesse ecossistema. Para Gorwa, ela envolve não apenas usuários e companhias de plataformas, como também, fundamentalmente, atores políticos de diferentes ramos do governo, outros *stakeholders* e grupos de *advocacy*, a exemplo de organizações não governamentais de privacidade e direitos digitais, da comunidade científica e de jornalistas investigativos.⁴

No entanto, é notório o papel desempenhado pelo setor privado, em especial diante do poder que intermediários possuem sobre a organização de suas próprias plataformas.

4 GORWA, Robert. What is platform governance? **Information, Communication & Society**, v. 22, n. 6, pp 854-871, 2019, p. 857. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1573914>. Acesso em: 24 ago. 2022.

Nesse meio, Gorwa destaca que a tendência são iniciativas de autorregulação com um único ator, que podem advir ou não de uma pressão governamental.⁵ Essas iniciativas, por sua vez, costumam ser traduzidas em políticas de comunidade ou termos de uso de plataformas, que contêm regras acerca de quais conteúdos são permitidos ou proibidos naquele espaço e que guiarão medidas como restrição de alcance e remoção de postagens de usuários.

Para Laura DeNardis, esse poder sobre o discurso online de que dispõem os intermediários pode ser considerado como uma “privatização da liberdade de expressão”.⁶ Seja em razão de pedidos governamentais por censura, como para aplicação da lei ou ganho político; seja em razão de preocupações com a reputação institucional ou valores e normas contratuais com o usuário final, em diferentes contextos são os regramentos privados que substituem (ou se somam) às leis, normas e governos no condicionamento da liberdade de expressão na esfera pública.⁷

Essa concentração de responsabilidades no ambiente digital - de determinar regras, administrá-las e executá-las - ressalta o poder que esses atores privados possuem e que impacta diretamente a vida das pessoas, sendo chamados, inclusive, de “novos governadores do discurso online”⁸ por alguns autores. Segundo Kadri e Klonick, além dos três poderes comuns ao Estado, as plataformas conteriam, ainda, uma quarta função, ao atuarem também como imprensa ou editores, isto é, “controlando o acesso ao discurso em nome de falantes e ouvintes”⁹.

Propostas na linha do Oversight Board, da Meta, surgem como alternativas para equilibrar essa balança, buscando envolver outros atores além das próprias companhias para influenciar no processo de moderação de conteúdo. No caso do Board, trata-se de um conselho de supervisão externa que advém de anos de pressão por instrumentos que possibilitassem a revisão das decisões de moderação

5 GORWA, Robert. The platform governance triangle: conceptualising the informal regulation of online content. **Internet Policy Review**, v. 8, n. 2, jun./2019. Disponível em: <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content>. Acesso em 30 ago. 2022.

6 DENARDIS, Laura. **The Global War for Internet Governance**. Yale University Press, 2014, p. 157.

7 *Ibidem*, p. 158.

8 KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. **Harvard Law Review**, v. 131, n. 6. p. 1598-1669, abr. 2018. Disponível em: <https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/> Acesso em: 30 ago. 2022.

9 “Platforms are both the governors, setting speech policies and adjudicating speech disputes, and the publishers, controlling access to speech on behalf of speakers and listeners”. In: KADRI, Thomas E.; KLONICK, Kate. Facebook V. sullivan: Public figures and newsworthiness in online speech. **Southern California Law Review**, v. 93, n. 37, pp. 37-99, 2019, p. 94. Disponível em: https://scholarship.law.stjohns.edu/cgi/viewcontent.cgi?article=1292&context=faculty_publications. Acesso em: 30 ago. 2022.

do Facebook.¹⁰ De acordo com Gorwa, porém, pode ser preciso encarar a criação desse conselho como mais um arranjo de governança privada, observando também que o discurso da empresa tem se afastado da ideia inicial de “Suprema Corte do Facebook”¹¹.

Na página do Oversight Board, o objetivo do conselho é descrito como

(...) promover a liberdade de expressão por meio da tomada de decisões independentes e baseadas em princípios com relação ao conteúdo no Facebook e no Instagram e por meio da emissão de recomendações sobre a política de conteúdo relevante da empresa do Facebook.

(...) O objetivo da criação do comitê não é ser uma simples extensão do processo de análise de conteúdo do Facebook que já existe. Em vez disso, ele analisará um número seletivo de casos altamente emblemáticos e determinará se as decisões foram tomadas de acordo com as políticas e valores declarados pelo Facebook.¹²

Seria possível verificar, em tese, uma espécie de substabelecimento ou divisão de poderes em relação à moderação de conteúdo em plataformas da Meta, especialmente no que tange à revisão de casos paradigmáticos e que servirão de base para recomendações gerais para a empresa. Em um projeto similar, o Twitter possui também o Conselho de Trust & Safety, formado por diferentes organizações e especialistas, que atua por meio de grupos de consultoria temáticos em áreas como “segurança e assédio online, direitos humanos e digitais, saúde mental e prevenção do suicídio, exploração sexual de menores e desumanização”, além da própria governança de conteúdo.¹³

Embora não se possa negar que o exercício da autorregulação contém vantagens específicas, como a possibilidade de realização de intervenções em determinadas

10 KADRI, Thomas E.; KLONICK, Kate. Facebook V. sullivan: Public figures and newsworthiness in online speech. **Southern California Law Review**, v. 93, n. 37, p. 37-99, 2019. Disponível em: https://scholarship.law.stjohns.edu/cgi/viewcontent.cgi?article=1292&context=faculty_publications. Acesso em: 30 ago. 2022.

11 GORWA, Robert. The platform governance triangle: conceptualising the informal regulation of online content. **Internet Policy Review**, v. 8, n. 2, jun./2019, p. 9. Disponível em: <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content>, Acesso em 30 ago. 2022.

12 COMITÊ DE SUPERVISÃO. **Comitê de Supervisão**. 2022. Disponível em: <https://oversightboard.com/>. Acesso em: 30 ago. 2022.

13 TWITTER. **Conselho de Trust & Safety**. 2022. Disponível em: <https://about.twitter.com/pt/our-priorities/healthy-conversations/trust-and-safety-council>. Acesso em 30 ago. 2022.

situações mais rapidamente do que por meio de legislações¹⁴, essa atuação encontra-se da mesma forma dependente da boa vontade de empresas privadas e muitas vezes não possui procedimentos transparentes ou explícitos o suficiente para permitir uma fiscalização efetiva por parte da sociedade civil.¹⁵ É por isso que atuação de outros atores pode servir para auxiliar no preenchimento dessas lacunas.

Ainda que não disponham de poder coercitivo, entidades do terceiro setor e da academia contribuem não apenas através da elaboração e produção de conteúdo científico sobre a matéria, como também por meio da fiscalização das atividades das plataformas e da participação em espaços de discussão que muitas vezes resultam em documentos de recomendações. Os Princípios de Santa Clara,¹⁶ assim como os Princípios de Manila¹⁷ são apenas alguns dos exemplos mais conhecidos de iniciativas que partiram da sociedade civil para colaborar com o cenário de moderação de conteúdo online.

Não obstante a ausência de caráter vinculante, é fato que essas recomendações auxiliam na promoção de um espaço virtual mais saudável e em respeito aos direitos humanos, seja a partir da influência no cenário legislativo, na pressão exercida para com o setor privado ou na criação de parâmetros de boas práticas. Em relação aos impactos da primeira versão dos Princípios de Santa Clara, por exemplo, a diretora de Liberdade Internacional de Expressão da Electronic Frontier Foundation, Jillian York, destacou ter percebido que empresas de tecnologia teriam melhorado, principalmente, em relação ao oferecimento de formas de contestação de decisões de moderação para usuários, o que demonstraria um avanço.¹⁸

No caso do setor governamental, esse auxílio se dá principalmente por meio de regulação, mas pode ocorrer, ainda, através do exercício de fiscalização, tanto pelo Poder Judiciário como pelo Poder Executivo. Na ausência de uma regulação mais específica, atualmente o cenário de moderação de conteúdo é bastante influenciado pelo Marco Civil da Internet, que apesar de não mencionar expressamente o termo

14 GORWA, Robert. What is platform governance? **Information, Communication & Society**, v. 22, n. 6, pp 854-871, 2019, p. 863. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1573914>. Acesso em: 24 ago. 2022.

15 GORWA, Robert. What is platform governance? **Information, Communication & Society**, v. 22, n. 6, pp 854-871, 2019, p. 863. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1573914>. Acesso em: 24 ago. 2022.

16 EFF – Electronic Frontier Foundation et al. **Santa Clara Principles on transparency and accountability in content moderation**. Disponível em: <https://santaclaraprinciples.org/pt/cfp/>. Acesso 20 dez. 2022.

17 EFF – Electronic Frontier Foundation et al. **Princípios de Manila sobre responsabilidade civil de intermediários**. Disponível em: <https://manilaprinciples.org/pt-br/principles.html>. Acesso 20 dez. 2022.

18 WHITELAW, Ben. Jillian C. York on the newly revised Santa Clara Principles. **Everything in Moderation**, 15 de dezembro de 2021. Disponível em: <https://www.everythinginmoderation.co/jillian-c-york-santa-clara-principles/>. Acesso em: 02 jan. 2023.

“moderação”, prevê a responsabilização dos intermediários pelo conteúdo publicado por terceiros somente no caso de descumprimento de ordem judicial própria, salvo exceções legais a alguns conteúdos, como ocorre com a disseminação não consensual de imagens íntimas. Desse modo, verifica-se que o regime de responsabilidade de intermediários regula indiretamente a matéria ao influenciar a estrutura de incentivos para a remoção ou não remoção de conteúdos.

A participação do Estado nessa esfera se traduz em um regime de moderação pautado pela correção, que seria aquele em que há um encorajamento ativo, apoio e eventual monitoramento da autorregulação por autoridades públicas.¹⁹ Nesse sentido,

Mecanismos de correção podem ter diferentes níveis de formalidade. Iniciativas correção, em um sentido estrito, geralmente incluem um elemento regulatório formal - como uma lei ou decisão administrativa - que age como uma estrutura e governa as atividades dos atores envolvidos, incluindo regras e consequências de diferentes tipos (tradução nossa).²⁰

A criação de uma norma estatal mais direcionada para o processo de moderação, porém, não é uma tarefa fácil, pois apresenta uma série de desafios. A lei alemã NetzDG, por exemplo, enfrentou esse problema e demonstrou uma série de dificuldades que podem ser geradas pela regulação governamental, como incentivos para que as plataformas digitais realizem a moderação de forma excessiva e a delegação ao setor privado da definição do que é legal e ilegal.²¹

Ainda assim, o papel do setor governamental permanece relevante ao estabelecer limites para essa atividade e assegurar garantias mínimas aos usuários. Em um contexto em que a maior parte das plataformas sociais advém de países do Norte global, o papel de reguladores nacionais pode contribuir para que essas empresas forneçam um serviço mais personalizado e atento também às particularidades dos países a que se destinam, como aqueles localizados no Sul global. Entretanto, escolher o regime de moderação mais adequado à realidade de cada Estado-nação é uma das grandes discussões no cenário atual.

19 ACCESS NOW. **26 recommendations of content governance**: a guide for lawmakers, regulators, and company policy makers. Disponível em: <https://www.accessnow.org/cms/assets/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf>. Acesso em: 19 dez. 2022.

20 ACCESS NOW. **26 recommendations of content governance**: a guide for lawmakers, regulators, and company policy makers. p. 11. Disponível em: <https://www.accessnow.org/cms/assets/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf>. Acesso em: 19 dez. 2022.

21 ESTARQUE, Marina; ARHEGAS, João Victor. **Redes sociais e moderação de conteúdo**: criando regras para o debate público a partir da esfera privada, Instituto de Tecnologia e Sociedade (ITS): Rio de Janeiro, 2021, p. 20. Disponível em: <https://itsrio.org/pt/publicacoes/redes-sociais-e-moderacao-de-conteudo/>. Acesso em: 19 dez. 2022.

Por fim, no âmbito dos modelos de regime de moderação, haveria ainda a possibilidade de adoção de um regime heterorregulatório, em que as normas seriam traçadas pelo Poder Público sem a participação das próprias plataformas. Essa modalidade, porém, enfrentaria dificuldades como, por exemplo, a falta de expertise do governo sobre a dinâmica de cada plataforma, o que poderia prejudicar a elaboração de normas alcançáveis ou coerentes ao seu funcionamento. No entanto, há que se ressaltar, no caso, a possibilidade de o próprio governo estabelecer comissões de especialistas ou chamadas de consultas públicas para auxiliar na elaboração de regulações, como ocorreu, por exemplo, no andamento do Projeto de Lei 21/2020, voltado a regular inteligência artificial no Brasil.²²²³

2.2. Modelo regulatório de moderação de conteúdo no Brasil

Conforme já mencionado, o Brasil carece atualmente de uma legislação mais específica acerca de moderação de conteúdo. O Marco Civil da Internet, aprovado há quase uma década, não precisou aprofundar a temática à época, considerando que os problemas mais urgentes, então, eram voltados a estabelecer direitos e garantias mínimas ao usuário da internet.

Apenas em seu art. 19, o legislador buscou eximir a responsabilidade de intermediários sobre o conteúdo postado por terceiros em suas plataformas, desde que não houvesse ordem judicial específica determinando alguma ação. Um estudo realizado por Hartmann e Monteiro apontou que esse foi o ponto que mais teve propostas de alteração por meio de projetos de lei desde a aprovação do MCI.

Segundo os autores,

Cerca de **33% dos projetos (16 PLs) propõem, em alguma medida, alterar o regime de responsabilização das plataformas** – seja eliminando por completo a dinâmica de notificação judicial estabelecida pelo art. 19, seja inserindo novas exceções a este regime, além dos casos de direito autoral e pornografia de vingança, já indicados no art. 19, § 2º e no art. 21.

(...) De modo geral, as principais preocupações dos parlamentares

22 AGÊNCIA SENADO. Inteligência Artificial: comissão de juristas entrega relatório nesta terça. 05 de dezembro de 2022. Disponível em: <https://www12.senado.leg.br/noticias/materias/2022/12/05/inteligencia-artificial-comissao-de-juristas-entrega-relatorio-nesta-terca>. Acesso em: 02 jan. 2023.

23 AGÊNCIA SENADO. Sugestões para lei sobre inteligência artificial podem ser enviadas até 13 de maio. 05 de abril de 2022. Disponível em: <https://www12.senado.leg.br/noticias/audios/2022/04/sugestoes-para-lei-sobre-inteligencia-artificial-podem-ser-enviadas-ate-13-de-maio>. Acesso em: 02 jan. 2023.

veiculadas nos projetos são a **propagação de notícias falsas, de atentados terroristas e de conteúdos ofensivos, que instiguem lesão corporal ou suicídio** (grifo nosso).²⁴

Assim, verifica-se que já há uma grande preocupação em atualizar o que foi previsto no MCI conforme as novas demandas que surgiram ao longo dos últimos anos, especialmente após episódios como as eleições presidenciais de 2018 e 2022. Diferentes pesquisas já surgiram para buscar compreender não apenas a natureza e característica desses fenômenos,²⁵ como também para avaliar possíveis limitações à disseminação desse tipo de conteúdo durante pleitos eleitorais.²⁶

É nesse contexto que surge, inclusive, o projeto de lei 2.630/2020, chamado “PL das Fake News”, que possuía inicialmente o objetivo de combater notícias falsas, mas a partir da criação de dispositivos que fomentavam a vigilância massiva de usuários, com a identificação de contas e coleta de registros de encaminhamento de mensagens.²⁷ Em que pese o objetivo de combater a desinformação ser legítimo, estudos chamaram atenção para os riscos à privacidade e direitos de personalidade dos usuários que regras nesse sentido poderiam trazer.²⁸

Ao longo da tramitação legislativa, porém, a proposta sofreu uma série de alterações, que acabaram tornando-a uma espécie de regulação geral voltada a plataformas digitais. A respeito de processo de moderação de conteúdo, o PL conta, por exemplo, com normas voltadas à construção de um devido processo para o procedimento de moderação e exigência de publicação de relatórios de transparência, além da criação do Conselho de Transparência e Responsabilidade na Internet, órgão voltado à fiscalização do cumprimento das regras pelas plataformas.

Apesar de não ser o foco, outras legislações, como a Lei Geral de Proteção de Dados e a

24 HARTMANN, Ivar; IUNES, Julia. Fake news no contexto de pandemia e emergência social: os deveres e responsabilidades das plataformas de redes sociais na moderação de conteúdo online entre a teoria e as proposições legislativas. **RDP**, Brasília, v. 17, n. 94, p. 388-414, jul./ago. 2020. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/download/4607/Hartmann%3B%20Iunes%2C%202020>. Acesso em 19 dez. 2022.

25 DOURADO, Tatiana Maria Silva Galvão. **Fake news na eleição presidencial de 2018 no Brasil**. 308 f. Tese (Doutorado) – Programa de Pós-Graduação em Comunicação e Culturas Contemporâneas, Universidade Federal da Bahia, Salvador, 2020.

26 SARLET, Ingo Wolfgang; SIQUEIRA, Andressa de Bittencourt. Liberdade de expressão e seus limites numa democracia: o caso das assim chamadas “faew news” nas redes sociais em período eleitoral no Brasil. **Revista Estudos Institucionais**, v. 6, n. 2, pp. 534-578, maio/ago. 2020. Disponível em: <https://www.estudosinstitucionais.com/REI/article/view/522/511>. Acesso em: 19 dez. 2022.

27 COALIZÃO DIREITOS NA REDE. **Nota técnica sobre o relatório de 26 de junho de 2020 ao Projeto Lei nº 2.630/2020**. 28 de junho de 2020. Disponível em: <https://direitosnarede.org.br/2020/06/28/nota-tecnica-sobre-o-relatorio-de-26-de-junho-de-2020-ao-projeto-lei-no-2-630-2020/>. Acesso em: 19 dez. 2022.

28 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: 19 dez. 2022.

construção de um Marco Legal para a regulação de inteligência artificial, em especial, por meio do projeto de lei 21/2020, que está em discussão no Senado Federal, poderão igualmente impactar na esfera da regulação de conteúdo online. Com relação a este último, apresentam maior destaque as regras voltadas à explicabilidade, transparência e responsabilização de sistemas de IA. Em um contexto marcado pelo amplo uso de sistemas de inteligência artificial para aplicação das normas de moderação, os aspectos supracitados de uma possível regulação sobre IA deverão impactar de forma significativa o tema da moderação de conteúdo.

Estas normas têm o condão de afetar diretamente o uso de tecnologias automatizadas para moderação de conteúdo, uma vez que tais sistemas podem apresentar problemas envolvendo viés algorítmico, discriminação e falhas técnicas em seu funcionamento. Ainda, considerando que essas ferramentas trabalham na limitação do discurso publicado online, atingindo também a liberdade de expressão dos indivíduos, despontam como essenciais normas que facilitem a compreensão e auditoria desses sistemas.

3. Metodologia

A presente seção se propõe a abordar a metodologia utilizada para o desenvolvimento da pesquisa. Assim, a divisão das etapas se deu nos seguintes tópicos: i) elaboração do roteiro de perguntas; ii) composição da amostra de pessoas entrevistadas; iii) procedimentos para a realização das entrevistas; e iv) categorias e códigos para análise de conteúdo das entrevistas.

Considerando a garantia do sigilo dos entrevistados, foram utilizados mecanismos de criptografia para resguardar a confidencialidade das informações durante todas as etapas da pesquisa. Para tal, as pesquisadoras fizeram uso dos softwares GpgEx e Kleopatra para a criação de chaves e cifragem dos arquivos, assim como de aplicativo de mensageria privada, com criptografia de ponta a ponta, para a transferência e partilha dos arquivos cifrados entre si.

3.1. Definição do caráter da entrevista e elaboração do roteiro de perguntas

Primeiramente, cumpre ressaltar que as autoras optaram por realizar uma entrevista de caráter semiestruturado, em que a elaboração do roteiro de perguntas serviu menos como um modelo rígido a ser seguido e mais como uma referência para direcionar o debate com os entrevistados. Assim, nem sempre os participantes eram obrigados a responder a todas as perguntas do roteiro. Essas variações se deram em razão da identificação de que determinadas questões já haviam sido respondidas de maneira espaçada ao longo da entrevista ou em virtude da trajetória do entrevistado,

para o qual a realização de determinada pergunta poderia não ser pertinente ao seu campo de atuação. Ainda, foi dada liberdade para que o próprio entrevistado optasse por responder ao questionamento ou passar para a próxima questão.

Dito isso, as principais etapas preliminares à elaboração do roteiro, que segue anexo no Apêndice I, foram a realização de reuniões com o consultor do projeto e com outros pesquisadores do IRIS, os quais já haviam realizado pesquisa também por meio de entrevistas (a saber, para a produção do *paper* “[Criptografia e Investigações Criminais: o que pensam os profissionais que movimentam o debate no Brasil](#)”)²⁹. Esses momentos, por sua vez, foram importantes para a definição de alguns procedimentos que foram utilizados nas fases seguintes, a exemplo da escolha da plataforma para a realização das entrevistas e escrita do Termo de Consentimento Livre e Esclarecido (TCLE).

Quanto à elaboração do roteiro, foram formuladas 18 perguntas, com base em pesquisa prévia realizada a partir de uma revisão sistemática de literatura estrangeira e nacional acerca dos diferentes regimes de moderação de conteúdo. Com duração aproximada de 50 minutos, o objetivo da entrevista era compreender as concepções dos entrevistados sobre os temas em questão. Assim, tais foram divididas em 3 blocos, quais sejam:

- I. *Apresentações e temas gerais*: 2 perguntas; objetivo de ser um momento de apresentação do entrevistado e melhor compreensão da sua trajetória em relação à moderação de conteúdo;
- II. *Moderação de conteúdo e a interação entre os diferentes atores envolvidos*: 5 perguntas; se propôs a explorar questões sobre o papel dos atores envolvidos nas práticas abordadas;
- III. *Modelos regulatórios*: 11 perguntas; foco nos modelos regulatórios e percepções sobre o Projeto de Lei nº 2630/2020.

29 Os autores do referido trabalho, Ana Bárbara Gomes e Gustavo Rodrigues, colaboraram com o compartilhamento de sugestões e contribuições ao roteiro da entrevista, assim como com relação a estratégias de criptografia e codificação das entrevistas, razão pela qual deixamos aqui registrado nosso agradecimento. Ainda, ao consultor metodológico do projeto, Lucas Caetano Pereira, agradecemos também pelos aportes materiais ao presente trabalho, desde a elaboração do roteiro até a fase final de redação, após as entrevistas.

3.2. Composição da amostra de pessoas entrevistadas

A fim de escolher os selecionados para compor a amostra de participantes, a busca teve início com a pesquisa por pessoas envolvidas com as discussões em torno do tema de moderação de conteúdo e seus subtemas. A busca contemplou espaços profissionais e acadêmicos, com destaque para atividades como autoria em trabalhos científicos, participação em eventos da área, contribuições jurídicas (tais como participação em audiências públicas referentes a projetos de lei relacionados aos temas) e engajamentos diversos.

As pessoas foram selecionadas buscando contemplar a participação dos quatro principais setores da governança da internet (a saber: setor governamental, setor privado, comunidade técnico científica e terceiro setor). No ponto, duas observações são importantes a respeito da amostra final: a) a primeira é que, dentro do setor governamental, foram consideradas pessoas atuantes não somente no Poder Executivo, como também em outras esferas atinentes ao setor público no Brasil de maneira geral; e b) além disso, apesar de as pessoas terem sido identificadas como pertencentes a um setor específico, havia casos em que o entrevistado poderia ser vinculado a mais de um setor, em virtude da sua atuação. Apesar disso, foi priorizada, ao longo do texto, a vinculação do participante ao setor pelo qual ele havia sido entrevistado, o que foi determinado de forma discricionária pelas pesquisadoras, conforme observação de sua atuação.

Neste ponto, é importante destacar que não se desconhece a dificuldade em estabelecer um conjunto objetivo de regras cientificamente válidas para determinar, no caso concreto, o pertencimento de pessoa X ou Y ao setor A ou B. Inclusive, há pesquisas sobre o caráter multissetorial da governança da internet que discutem a ficcionalidade dessa divisão entre setores. Gustavo Rodrigues refere-o, com base no conceito de ciborgue da filósofa Donna Haraway, como “uma ficção material e semiótica simultaneamente descritiva e normativa, real e imaginada”³⁰. Para Hofmann,

A difícil correspondência entre a taxonomia das partes interessadas e o espectro político na governança da Internet está em desacordo com a ideia básica de multissetorialismo, que pressupõe que as posições políticas podem ser agregadas de acordo com as afiliações formais (tradução nossa).³¹

30 RODRIGUES, Gustavo Ramos. “A função de uma lei de proteção de dados é proteger a todos, inclusive aquele que coleta dados”: o debate sobre a Lei Geral de Proteção de Dados e a ideologia da harmonia no Fórum da Internet no Brasil. In: VII ENCONTRO NACIONAL DE ANTROPOLOGIA DO DIREITO, 2022, São Paulo. **Anais [...]**. São Paulo: NADIR/USP, 2021. v. 1. p. 12. Disponível em: bit.ly/3GBIZYx. Acesso em: 11 jan. 2023.

31 “The difficult match between the stakeholder taxonomy and the political spectrum in Internet

Ainda assim, a escolha por determinar o setor do entrevistado conforme observação de sua atuação e afiliação, mesmo diante de eventual participação em múltiplos setores, serviu enquanto instrumento metodológico para análise de um campo social amplamente estruturado no plano êmico pela ideia de multissetorialismo. Nesse sentido, importa destacar o caráter heurístico dessa divisão setorial, sem buscar reificá-la como algo estático, fixado e simplesmente transponível a outros contextos nos quais outros esquemas classificatórios podem operar.

Os convites, por sua vez, foram então enviados por e-mail, divididos entre as pesquisadoras. O Termo de Consentimento Livre e Esclarecido foi enviado via plataforma Docusign, após o aceite, e o roteiro poderia ser enviado mediante solicitação do convidado. O TCLE Palcontinha informações quanto ao objetivo da pesquisa, justificativa, procedimentos, medidas de sigilo e privacidade, dentre outros. Foram concluídas 18 entrevistas entre os meses de maio e julho, e, em decorrência da saída de uma das pesquisadoras, mais uma entrevista foi realizada no mês de dezembro, a fim de complementar o quadro de entrevistados. Da amostra final, 4 pessoas eram representantes do setor governamental, 3 do setor privado, 5 da comunidade técnico-científica e 6 do terceiro setor (conforme definição mencionada anteriormente). Para manter o sigilo dos participantes, foram gerados pseudônimos para cada um, de maneira aleatória e por meio de plataforma online.

3.3. Procedimentos para a realização das entrevistas

As entrevistas foram, preponderantemente, realizadas pela plataforma Zoom. Dentre as 2 exceções, 2 foram excepcionalmente enviadas por escrito, visto que os entrevistados preferiram esse método de participação devido a conflitos de agenda. Ainda, uma das entrevistas escritas foi elaborada pela equipe de trabalho do entrevistado convidado.

Quanto às que foram realizadas por meio de videochamada, o consentimento para gravação foi colhido através do TCLE, não havendo nenhuma negativa nesse sentido. Visto o caráter semiestruturado da entrevista, o roteiro foi utilizado como referência, não havendo rigidez quanto à ordem das perguntas e com liberdade quanto à realização de perguntas voltadas a aprofundar o tema, a fim de melhor compreender as respostas dos entrevistados.

governance is at odds with the basic idea of multi-stakeholderism, which assumes that political positions can be aggregated along the lines of formal affiliations". In: HOFMANN, Jeanette. Multi-stakeholderism in Internet governance: putting a fiction into practice. *Journal of Cyber Policy*, v.1, n. 1, pp. 29-49. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/23738871.2016.1158303>. Acesso em: 11 jan. 2023.

As entrevistas foram transcritas pelas pesquisadoras e pelo consultor. Foi utilizado, para auxílio dessas atividades, o programa InqScribe, o qual permite a execução do áudio/vídeo em conjunto à transcrição, mas sem coletar dados do conteúdo transcrito.

3.4. Categorias e códigos para análise de conteúdo das entrevistas

Para a fase de análise de conteúdo das entrevistas, foram realizadas codificações através do Atlas.ti 8, um programa que “contribui para a organização dos dados, a otimização do processo analítico, além de ser maleável para inúmeros tipos de pesquisas qualitativas com objetivos diversos”³². Assim, o primeiro passo foi a criação dos códigos que seriam utilizados, etapa que foi realizada em conjunto pelas pesquisadoras, em prol de sintetizar as categorias de maior relevância e padronizar as informações que seriam coletadas a partir do roteiro. A análise qualitativa sistemática em questão apresentou caráter indutivo, uma vez que não se partiu de um conjunto pré-definido de critérios para criação dos códigos e categorias utilizados.

Dessa maneira, foram criados 83 códigos (esquema de codificação no Apêndice II), divididos entre as pesquisadoras, de forma que cada uma realizou a codificação a partir de um conjunto de códigos. Posteriormente, foram geradas planilhas pelo programa, em formato Excel, com o resultado das codificações. Tais foram condensadas em uma única planilha para análise, a qual foi compartilhada entre as pesquisadoras de maneira criptografada. A partir desse material, foram analisados os conteúdos advindos de cada pergunta do roteiro e escritos os tópicos que formam o próximo capítulo deste trabalho, através de uma abordagem interpretativa e da recriação narrativa dos argumentos encontrados.

3.5. Limitações da pesquisa

Antes de partir para a análise dos resultados, cabe ressaltar que a presente pesquisa possui algumas limitações. A primeira delas é que o presente trabalho não apresenta resultados representativos de determinado setor da governança da internet ou de qualquer outro segmento populacional. Isto se deu em razão do caráter assistemático do método de seleção de amostra, que não produziu um universo empírico representativo das populações que se pretende analisar.

32 SILVA JUNIOR, L. A.; LEAO, M. B. C. O software Atlas.ti como recurso para a análise de conteúdo: analisando a robótica no Ensino de Ciências em teses brasileiras. **Ciência & Educação**, Bauru, v. 24, n. 3, p. 715-728, set. 2018. Disponível em: <https://www.scielo.br/j/ciedu/a/yBwC9L74v4vD3s4PwVXggsk/?lang=pt>. Acesso em: 21 dez. 2022.

Nesse sentido, também cabe destacar a dificuldade enfrentada em realizar uma quantidade igualitária de entrevistas entre os diferentes setores, seja por meio de contato ou disponibilidade de agenda, de forma que o setor técnico-científico e terceiro setor encontram-se sobrerrepresentados em comparação aos demais. Isto pode representar um viés em relação às conclusões obtidas ao final.

Por fim, o objetivo deste relatório foi desenhar um panorama em torno das principais opiniões que circundam o tema de moderação de conteúdo entre os atores especializados envolvidos, com o objetivo de contribuir para o debate público em torno do assunto.

4. Resultados: análise das entrevistas

Neste capítulo, são apresentados os resultados encontrados na pesquisa, a partir da análise e discussão dos argumentos apresentados pelos entrevistados dentro das diferentes temáticas.

Os argumentos foram analisados por meio de reconstrução narrativa, buscando apresentar as relações lógicas existentes entre eles e com destaque para as afirmações mais frequentes. Em conjunto, são tecidos comentários, acompanhados de citações a marcos teóricos, a fim de debater os enunciados.

Ainda, ao final de alguns tópicos são apresentados mapas mentais, a fim de ilustrar os principais argumentos levantados pelos entrevistados.

4.1. Sobre moderação de conteúdo e suas técnicas

Os argumentos a seguir foram encontrados a partir das respostas dos entrevistados às perguntas 3 e 13 do roteiro, que abordavam sobre as principais funções e atividades da moderação de conteúdo, assim como da avaliação em torno do uso de mecanismos automatizados (inteligência artificial) para essa atividade e a necessidade de revisão humana.

Sobre as principais funções e ações da moderação de conteúdo

De forma introdutória, as primeiras perguntas do roteiro - com caráter conteudista - se propuseram a abordar quais seriam, para os entrevistados, as principais funções da moderação de conteúdo e quais ações estariam contempladas em seu escopo de atuação. Visto a grande correspondência que as respostas apresentaram entre si, foi possível dividi-las em grandes grupos.

Assim, dentre as funções elencadas, a divisão foi entre: **i. Sanitização do espaço virtual (argumento mencionado 14x)** e **ii. Produto das plataformas (8x)**. Já quanto às ações entendidas como moderação de conteúdo, a divisão foi entre: **i. Curadoria de conteúdo (10x)** e **ii. Intervenção (13x)**. Assim, visto a presença de pontos em comum, mas também de questões distintas e particulares nas concepções dos entrevistados, torna-se válido discorrer sobre cada um dos tópicos de maneira individual.

Função: Sanitização do espaço virtual

Plataformas digitais podem atender a fins diversos. Embora atualmente as grandes redes sociais tenham apresentado sucesso em servir aos mais variados interesses em uma só plataforma, a *web* é composta por uma miríade de portais e comunidades. Por consequência, padrões de comportamento precisam ser definidos para garantir uma navegação segura e harmônica aos usuários, além de restringir ou mesmo impedir atos considerados inapropriados. Nesse sentido, apesar das grandes plataformas apresentarem membros de diversas nacionalidades, etnias e credos, algumas regras de convivência são universais para que a experiência dos usuários seja possível e, além disso, satisfatória.

Partindo desse pressuposto, um ponto apresentado foi que a promoção do caráter “saudável” do espaço virtual cabe às práticas de moderação de conteúdo. Nessa perspectiva, o ato de “sanitizar” a rede poderia ser definido como a obstrução de conteúdos desagradáveis, nocivos ou mesmo ilícitos, tendo como principais exemplos os discursos de ódio, pornografia infantil, terrorismo, racismo, desinformação, spam, violência online e atos como cyberbullying e doxing. Nesse diapasão, a garantia desse filtro básico foi referenciada por meio de expressões como: determinação de um mínimo civilizatório, manutenção de um espaço receptível aos usuários e materializar um ambiente minimamente habitável, propício à conversa ou passível de interação.

Em conjunto à obstrução de tais conteúdos, alguns entrevistados também atrelaram ao ato de “sanitizar” o espaço virtual a **garantia da segurança (3x)**, **a participação civil (1x)** e **a liberdade de expressão dos usuários (2x)**. Nesse contexto, Maria Teixeira, membra da comunidade técnico-científica, afirma:

Eu acho que a função primordial da moderação de conteúdo seria **evitar ou então constranger atividades criminosas de uma maneira geral**. Então crimes em todos os sentidos: crimes de racismo, enfim, crimes de aliciamento de menores, esse tipo de coisa que é mais fiscalizado, digamos assim. Mas as funções paralelas dessa moderação de conteúdo são justamente **identificar como esses conteúdos não são somente criminosos nesse sentido, são mais conteúdos que parecem inofensivos,**

mas que por trás deles existem muitas outras coisas. São conteúdos mentirosos, são conteúdos que podem destruir a imagem pública de determinados indivíduos. Então eu acho que a função da moderação de conteúdo, uma das funções, uma função paralela além de constranger crimes, seria **constranger a extrapolação da liberdade de expressão.** Então eu acho que a moderação de conteúdo é isso. Primeiro, **evitar atividades criminosas e segundo, evitar esses discursos danosos de uma maneira geral.** Que nem sempre induzem a crimes, mas que são uma extrapolação da liberdade de expressão (grifo nosso).

Observa-se assim através da fala da entrevistada a relação compreendida entre a sanitização da rede e o estabelecimento de parâmetros para a liberdade de expressão dos usuários, ao passo que o ato de evitar discursos com caráter danoso configura-se como uma maneira de garantir um espaço respeitoso.

Função: Produto das plataformas

A moderação de conteúdo é uma prática que, além de servir aos usuários, serve às plataformas, tornando possível a execução de seu objeto e, logo, garantindo sua utilidade. Nesse contexto, a função de produto pode ser definida como um mecanismo da plataforma que assegura o cumprimento das regras internas em prol do modelo de negócios vigente (quando há um caráter comercial) ou simplesmente em acordo ao cumprimento de sua proposta inicial. Visto isto, independentemente do tipo da plataforma, a moderação é aqui compreendida como uma maneira de assegurar seu bom funcionamento. Quando relacionada a interesses comerciais, por sua vez, são contemplados aspectos de marketing, relacionados ao direcionamento de propagandas aos usuários, conforme citado por Fernanda Rezende, entrevistada membra do terceiro setor. No entanto, é mister ressaltar que não são necessários interesses econômicos para que as plataformas sintam necessidade de preservar suas regras através da moderação de conteúdo. Caio Fernandes, membro do governo, ilustra isso em sua fala quando diz que

(...) a plataforma tem todo o direito de ter suas regras e tem todo o direito de suspender o usuário que não está de acordo com as regras. Se você tem um clube que é de dança de salão, você não pode entrar de sandália havaiana porque não dá para fazer dança de salão com sandália havaiana. Por outro lado, se for um clube de nudistas, você não pode entrar vestido, são regras do clube. Isso não quer dizer que tenha um crime envolvido. Então, porque essas coisas às vezes se misturam.

Nesse contexto, a fala de Caio exemplifica também como a função de produto, por sua vez, difere das já expostas, ao passo que visa a assegurar não um filtro básico,

como no caso da sanitização ou o cumprimento de regras e padrões externos, mas sim a adequação às regras internas da plataforma - que podem servir a interesses comerciais ou não. Nesse contexto, Theo Freitas, membro do terceiro setor, exemplifica essa concepção da seguinte forma:

(...) é uma função temática. Então, a moderação de conteúdo, pensando nela como uma atividade que é **a atividade de elaboração e implementação de regras, padrões, normas de comunidade em setores de internet abertos à atividade de usuários**, abertos à geração de conteúdo por terceiros, pensando nisso, eu acho que a primeira função é uma função temática (grifo nosso).

Em um sentido semelhante, Samuel Cardoso, membro da comunidade técnico-científica, afirma que a principal função vai ser o manter o funcionamento adequado da plataforma, o qual vai variar conforme a política interna, objetivo e noção do que é “adequado” para a plataforma.

Ainda, 3 entrevistados associam à função de Produto a aplicação de padrões e diretrizes legais presentes no ordenamento jurídico no qual a plataforma está inserida. Esse complemento, por sua vez, se aproxima da função de Sanitização; no entanto, não se enquadra em tal, visto que os padrões mencionados pelos entrevistados atendem a regras mais gerais presentes nos respectivos ordenamentos. Citando novamente Samuel, o entrevistado comenta sobre o caráter regulatório, em diálogo à função de produto, ao complementar sua fala anterior da seguinte maneira:

(...) nessa ideia mais ampla de adequação, eu estou considerando que entra também **uma adequação à legislação, uma adequação à promoção de direitos humanos, uma adequação a padrões éticos**, eu entendo que a moderação tem essa função, de manter adequada à função, aos interesses da plataforma e aí vai variar, é, indo no limite, assim, uma plataforma de publicação de pornografia, por exemplo, poderia remover um filme do Dr. Estranho que tivesse publicado na plataforma e que não tem conteúdo adulto, né? Então, também eu chamaria isso de moderação, não só pela violação de direito autoral, para coibir, mas também para adequar o conteúdo da plataforma ao seu objetivo (grifo nosso).

Ao falar sobre moderação de conteúdo, grande parte das pessoas pensa logo em remoção, como se tais termos fossem sinônimos. Embora remover conteúdos na rede seja um dos atos mais controversos e, talvez por isso, um dos mais lembrados, moderar contempla uma série de atividades. Dentre elas, algumas citadas pelos entrevistados

foram: redução de visibilidade (ou downranking), rotulação, shadowbanning³³, suspensão, ranqueamento, invisibilização, banimento e verificação.

Nesse contexto, cabe ressaltar que os termos Intervenção e Curadoria não alcançaram uma distinção exata em seus entendimentos. Visto isto, a categoria “Moderação” é ora percebida como sinônimo de remoção, ora em sentido mais expansivo.

Ação: Intervenção

Conforme introduzido, a intervenção é aqui compreendida como atos de moderação de conteúdo com ação e efeito automáticos. Nesse cenário, os entrevistados associaram tais atividades, frequentemente, à execução de uma regra de plataforma que proíbe determinados tipos de conteúdos ou ações em seu ambiente. Assim, Theo Freitas, membro do terceiro setor, menciona atos de intervenção relacionados à aplicação de regras internas da plataforma e, logo, à função de Produto:

É... por exemplo, **ter regras que vão dizer ‘esse tema não pode aqui’**, então, exemplo: uma plataforma que é de resenha de filme. E tem lá um negócio que não é de resenha de filme, então isso... resolve, digamos assim, criar uma regra, implementar uma regra pra isso. É... ela garante uma multiplicidade de espaços e comunidades diferentes entre si, que vão poder discutir temas diferentes e vão poder, eventualmente, **excluir assuntos que não tem uma pertinência com aquilo** (grifo nosso).

Frente aos atos de intervenção contemplados na moderação de conteúdo, alguns entrevistados mencionaram as questões polêmicas que podem vir à tona com sua execução, tal como a remoção de conteúdos legais e adequados às regras da plataforma de maneira equivocada. Nesse âmbito, um exemplo recorrente é em relação aos casos de nudez, em que as plataformas removem automaticamente imagens que consideram inapropriadas, não levando em consideração aspectos culturais, artísticos, geracionais, entre outros, que atribuem diferentes sentidos à sua manifestação. Como exemplo, alguns entrevistados citaram o caso da fotografia intitulada *Napalm Girl*, a qual pode ser compreendida como conteúdo de nudez (e,

33 Em tradução livre, “banimento fantasma” ou “banimento pela sombra”. Trata-se de uma significativa redução de alcance nos conteúdos publicados por um usuário, como se tivesse ocorrido um banimento, mas sem que haja qualquer notificação por parte da plataforma. Para Radsch, “Shadowbanning” é comumente uma resposta a interações tóxicas ou indesejáveis relacionadas a um usuário específico. As contas que são ‘shadowbanned’ geralmente parecem estar funcionando normalmente para o usuário afetado, mas o seu conteúdo é indetectável e pode ser incapaz de se envolver de certas maneiras” (tradução nossa). In: RADSCH, Courtney. Shadowban/Shadow Banning. In: BELLI, Luca; ZINGALES, Nicolo; CURZI, Yasmin (orgs.). Glossary of platform: law and policy terms. Rio de Janeiro: FGV Direito Rio, 2021. Apesar de o objetivo ser a retirada de conteúdo potencialmente danoso das plataformas, essa prática afeta o direito de liberdade de expressão dos usuários sem lhes dar a possibilidade de contestação, assim como de ampla defesa e contraditório, uma vez que não são notificados acerca da restrição imposta.

logo, removida de maneira automática pelas plataformas); no entanto, trata-se de uma imagem com caráter histórico, visto que registra um bombardeio realizado durante a Guerra do Vietnã. Como os atos de intervenção realizados pelas plataformas são, preponderantemente, efetuados por meio de inteligência artificial, conteúdos adequados, muitas vezes, são interpretados erroneamente, visto as limitações da IA (assunto que será melhor explorado no tópico seguinte).

Ação: Curadoria de conteúdo

A curadoria de conteúdo se destaca por contemplar atividades não apenas com efeitos instantâneos, mas que muitas vezes buscam por modelar os comportamentos dos usuários e/ou ranquear conteúdos, visando assim efeitos mais complexos e a longo prazo. Por vezes foi colocada pelos entrevistados como um tipo de ação oposto à intervenção, já em outros casos, foi estabelecida uma relação de gênero e subgênero.

Dito isso, tem-se que a curadoria de conteúdo pode atender tanto à função de sanitização quanto a de produto das plataformas, embora esteja mais relacionada à segunda, visto que pode se atrelar a interesses distintos, sejam eles o de atender às regras internas da plataforma ou ao modelo de negócios vigente.

Nesse sentido, Samuel Cardoso, membro da comunidade técnico científica, discorre sobre algumas formas de atuação da Curadoria:

(...) no meu entender, a moderação de conteúdo abrange **desde a valorização e o ranqueamento de conteúdos por meio da limitação da divulgação ou do incentivo à divulgação e da ampliação desse acesso a determinado conteúdo ou da restrição de determinado acesso**. Então, sei lá, quando a foto de mudança do perfil ou o parabéns alcança mais **amplitude de difusão** do que uma mensagem triste, eu entendo que isso entra na moderação de conteúdo, porque a plataforma entende que isso é mais **adequado**, no caso, para manter as pessoas **acessando e utilizando os serviços da plataforma** (grifo nosso).

Nesse sentido, a adequação de conteúdos à plataforma também foi apontada como sendo realizada pela curadoria quando diz respeito ao caráter regulatório. Nesse sentido, Igor Peixoto, membro do setor empresarial, afirma:

(...) mas é claro que também a partir da moderação de conteúdo, quando as plataformas estão **fazendo essas regras** e analisando os conteúdos e removendo certos conteúdos, **diminuindo a visibilidade**, porque tem várias formas da moderação de conteúdo de incidir no conteúdo, elas também acabam, por tabela, de certa forma, **reduzindo a incidência de materiais**

ilícitos, que é algo que pode gerar uma **responsabilização** - e aí quando eu digo ilícito, e aí **depende do Estado, porque elas operam em escala global, então em cada Estado existe um ordenamento próprio**, existem tipos de conteúdo diferente que são considerados ilícitos, então é meio como se elas já fossem **fazer uma filtragem daquilo que possa ser considerado ilícito**. E até uma prática comum das plataformas, quando elas recebem comunicação do Estado, de representantes seja do Judiciário, seja, enfim, do Executivo, pedido de remoção de conteúdo ou de análise de conteúdo, elas primeiro analisam se aquilo lá viola alguma das políticas da plataforma, que aí se violar, elas removem e dizem que, enfim, “ah, violou as normas da plataforma” e não necessariamente as normas do Estado e tal e aí essa análise acaba ficando prejudicada. (...) então, de certa forma também é um meio de **umentar o nível de aderência às legislações locais** (grifo nosso).

A menção que Igor realiza à redução de conteúdos ilícitos por ações de curadoria se relaciona à fala de Bárbara Silveira, membra do setor empresarial, ao afirmar que

A restrição **de alcance de um conteúdo** permite uma abordagem mais sutil para os tipos de **discurso que podem ser considerados indesejados**, atingindo um melhor equilíbrio entre liberdade de expressão e liberdade de informação. Enfim, como a atenção é direcionada é uma questão-chave (grifo nosso).

No que diz respeito à modulação de comportamentos por curadoria de conteúdo, esse tópico muito se relaciona a questões emblemáticas debatidas no campo, tais como a polarização das redes e a veiculação de desinformação e fake news por filtros bolha. Assim, ao segmentar públicos e afetar o alcance dos conteúdos nas plataformas, os usuários passam a ser afetados de maneiras diversas, influenciando na esfera social e política. Por outro lado, as plataformas se beneficiam ao alcançarem maior engajamento em conteúdos publicitários e/ou um maior tempo de uso pelos indivíduos, o que, por sua vez, intensifica a obtenção de lucro.

Nesse sentido, Lucas Oliveira, membro do terceiro setor, discorre sobre como ações de Curadoria “moderam” comportamentos no seguinte trecho:

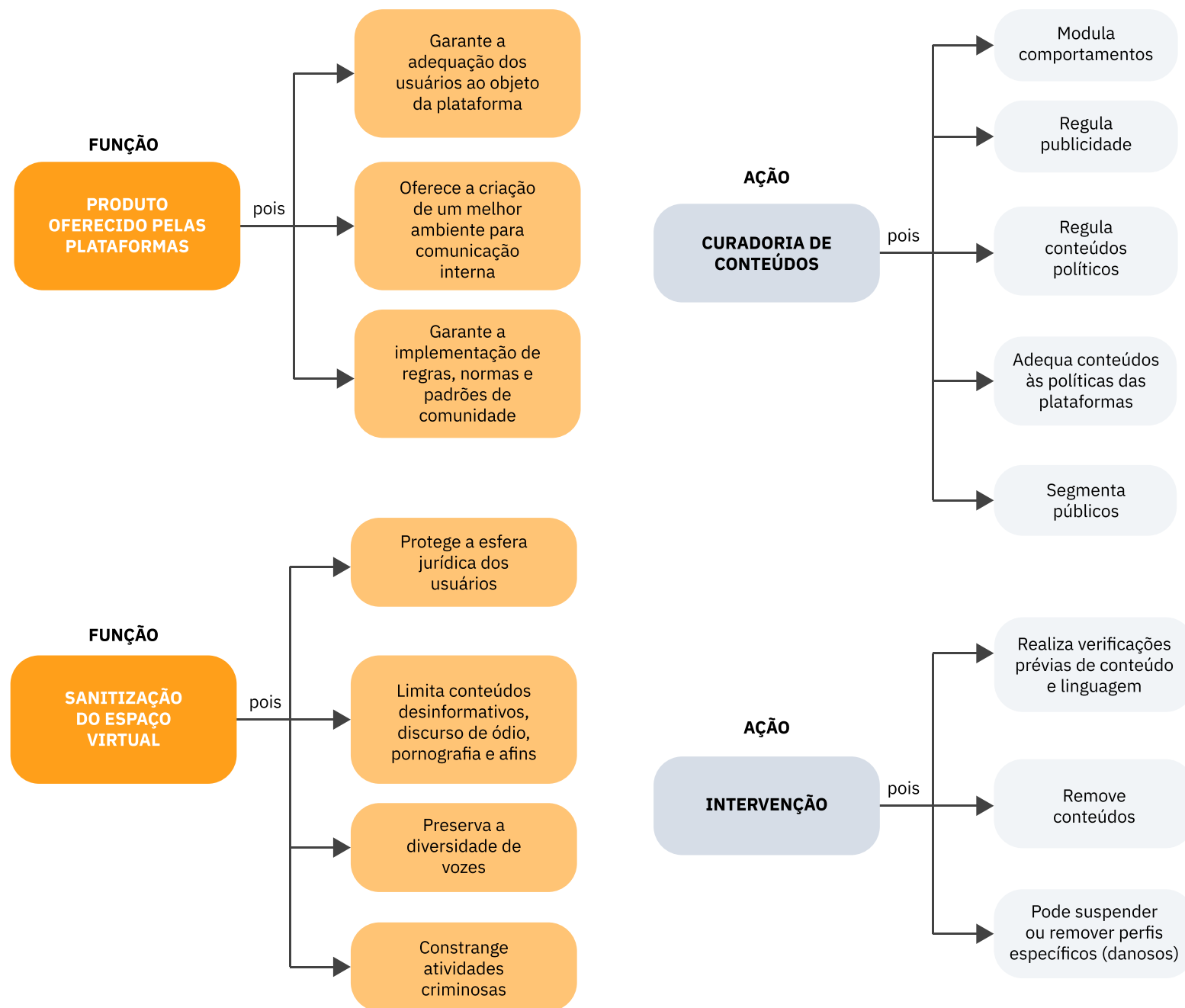
Olha, eu acho que moderação de conteúdo é **desde você remover uma mensagem, uma imagem de dentro de uma plataforma digital até você eventualmente limitar a distribuição daquela mensagem ou escolher segmentar quais públicos vão receber**

aquela determinada mensagem e hoje em dia pode expandir até um pouco mais e falar não só de moderação de conteúdo, como também de **moderação de comportamento**, a maneira como as pessoas se relacionam com essas novas tecnologias. Então você **colocar um rótulo** numa mensagem pra tentar evitar que as pessoas acreditem numa mensagem que é distorcida, ao invés de remover aquela mensagem da plataforma, às vezes tem muitas mais a ver com uma **moderação de comportamento, do que de conteúdo propriamente dito** (grifo nosso).

Por fim, tem-se que o caráter mais complexo das ações consideradas curadoria muito se atrela ao fato de tais serem realizadas apenas por meio de Inteligência Artificial e nem sempre haver transparência a respeito das diretrizes e métodos utilizados pelas plataformas. Alguns entrevistados citaram esse aspecto como sendo uma questão problemática, ressaltando a necessidade das diretrizes de moderação de conteúdo serem citadas e exemplificadas nos relatórios de transparência divulgados pelas plataformas.

Confira abaixo o mapa mental sintetizando os principais argumentos apresentados em relação às ações e funções da moderação de conteúdo:

Figura 1 - Principais argumentos sobre ações e funções da moderação de conteúdo



Sobre o uso de inteligência artificial para moderação de conteúdo

O uso de inteligência artificial para fins de moderação de conteúdo já é uma realidade para boa parte das plataformas digitais. Um dos principais argumentos favoráveis à sua utilização foi justamente o fato de que pelo tamanho da maioria das plataformas, que podem ter até milhões de usuários e, conseqüentemente, uma grande quantidade de postagens por hora, **a escala de conteúdo produzido (9x)** não permitiria que a moderação fosse feita somente por humanos.

Em temas como pornografia infantil, terrorismo, nudez e direito autoral, por exemplo, reconheceu-se que a **moderação automatizada teria alto grau de eficiência (3x)**, tendo como um dos motivos o compartilhamento de bases de dados entre as empresas. Entretanto, um dos problemas mais apontados foi a **falha que essas tecnologias apresentam na análise de contexto e/ou linguagem (7x)**. O posicionamento de Samuel Cardoso, membro do setor técnico-científico, ilustra bem essa preocupação e possíveis conseqüências:

Eu acho que, como a gente está falando de plataformas muito pequenas ou muito grandes, a questão é contextual e **os mecanismos automatizados, eles erram justamente na detecção dos contextos em sintonia fina**. Então, eu me preocupo muito com a porcentagem baixa sendo legitimada como um grau de eficiência desses mecanismos, porque quando a gente fala de milhões de comunicações, **essa porcentagem baixa pode significar centenas de milhares de erros, sejam de falsos negativos, mas principalmente como falsos positivos**. Então, eu acho que a remoção automatizada ou qualquer tipo de moderação automatizada **deveria ser vista mais como um procedimento precário, no sentido de algo que possa ser contornado mais rapidamente a partir de uma requisição**, a partir eventualmente de própria denúncia ou aviso pelos próprios usuários. (grifo nosso).

De fato, a **possibilidade de impugnação ou apresentação de recurso de decisões automatizadas (2x)**, diversificação de canais de contestação, assim como a **necessidade de assegurar uma revisão humana (4x)** apareceram como alternativas para mitigar os riscos dessas ferramentas. Para Emanuella Nogueira, membro do Terceiro Setor, é preciso se preocupar com o uso correto da moderação automatizada mais do que se devemos utilizá-la ou não:

(...) me parece que a gente precisa avançar qual é o uso correto, adequado desses mecanismos mais do que, enfim, se eles devem ser usados ou não, porque **eles têm uma série de problemas e**

o jeito de lidar com isso é pensar nesses processos e no que deve ser feito pra aprimorar a relação dele também com os humanos, digamos assim. (grifo nosso).

Nesse sentido, a entrevistada aponta a necessidade de realização de avaliação de impacto em direitos humanos, preocupações em torno de diversidade e com o processo de desenvolvimento e acompanhamento desses sistemas. Além disso, **menciona também a necessidade de maior transparência, somando-se a outros 4 entrevistados**, dentre os quais defendeu-se, ainda, o fornecimento de informações precisas sobre os critérios e procedimentos de decisões automatizadas, seu grau de precisão e maior explicabilidade aos usuários, até mesmo para ganhar a sua confiança.

Tal solução vai ao encontro de um dos argumentos negativos apresentados em torno do uso de IA, que alega haver pouca informação sobre os motivos de uma eventual moderação e sobre o processo de revisão. Considerando que **a possibilidade de existência de vieses nesses sistemas foi igualmente ventilada pelos participantes (2x)**, como, por exemplo, em razão do uso de bases de dados já enviesadas ou por quem projeta o algoritmo, é possível concluir que uma maior transparência também poderia auxiliar nesses casos.

Outras iniciativas também foram mencionadas, como uma moderação de conteúdo automatizada auditável, com critérios antidiscriminatórios explícitos, **times de moderação melhor instruídos acerca de contextos e especificidades regionais (2x)**, realização de avaliação de risco e a utilização de inteligência artificial para fins de automatização dos processos de denúncia e avaliação de denúncia de conteúdo. Para Igor Peixoto, membro do setor empresarial, é interessante priorizar o uso de IA para as situações em que há elevado grau de certeza de que determinado conteúdo viola as políticas de comunidade.

Segundo o entrevistado, um dos problemas da moderação automatizada é a crença de que a tecnologia seria a solução para todos os problemas:

Também o que eu vejo que eu já mencionei é essa **percepção de que a inteligência artificial soluciona todos os problemas**. “Ah, eles podem desenvolver uma tecnologia que faz isso”, não é tão simples assim e a gente sabe que as tecnologias, como eu mencionei, elas cometem muitos erros, então não dá pra gente ficar completamente dependente delas (grifo nosso).

Caio Fernandes, membro do setor governamental, reforça essa percepção ao afirmar que considera que os mecanismos automáticos para moderação ainda são imaturos. Para Luiz Porto, em alguns casos não há que se falar, inclusive, em

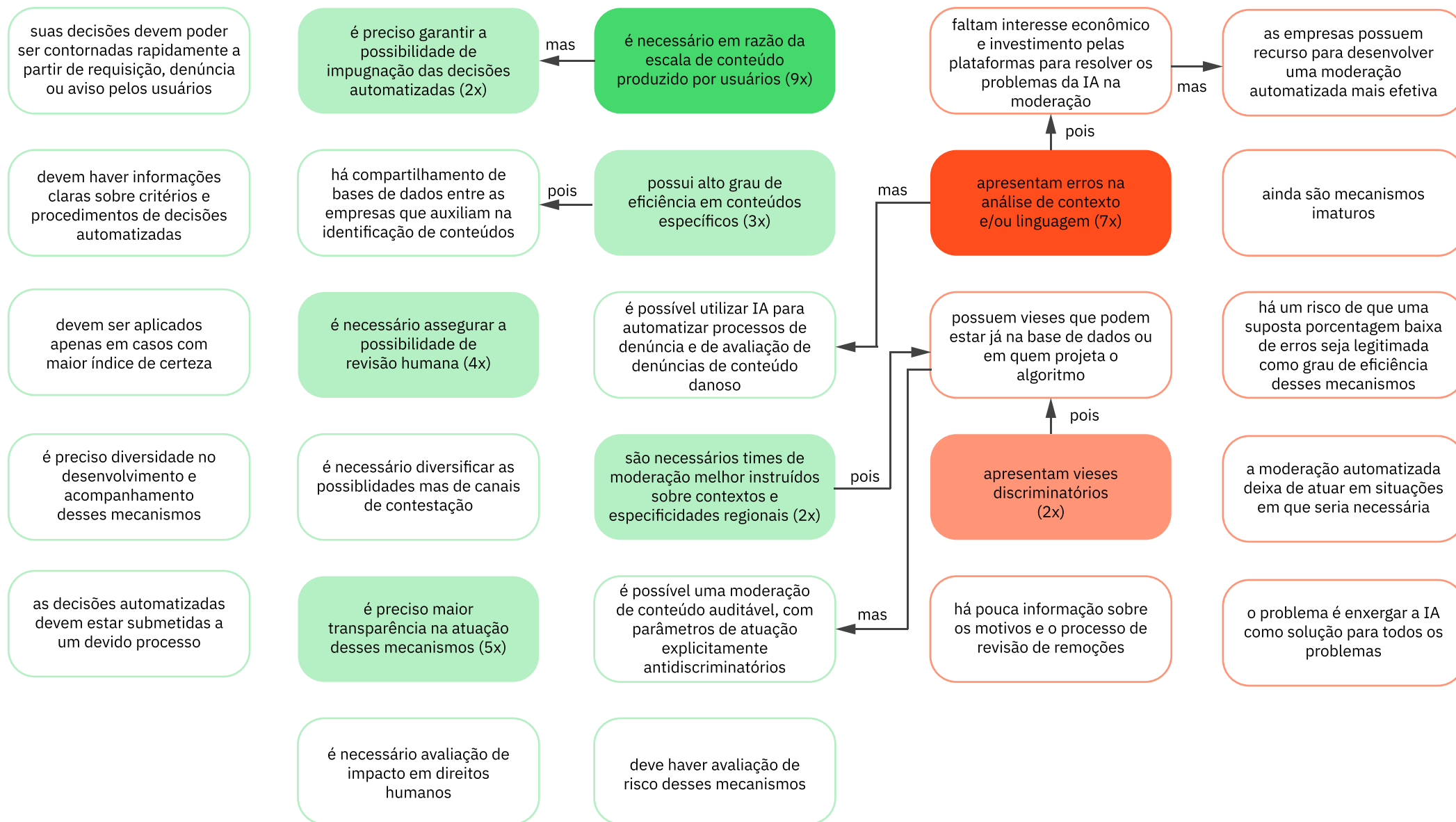


decisões contrárias a exercícios de direitos e liberdades fundamentais, mas de uma moderação automatizada que deixou de atuar em situações em que seria necessária.

Para o entrevistado, é preciso, ainda, garantir que decisões automatizadas “acompanhem o justo e devido processo e que por meio do uso da inteligência artificial, a própria inteligência artificial esteja submetida aos crivos do devido processo do direito como são estabelecidos”. Entretanto, há de se considerar o apontamento feito por outro participante, no sentido de que, atualmente, haveria pouco investimento e interesse econômico por parte das plataformas em solucionar problemas relacionados à moderação de conteúdo por IA e torná-la mais efetiva, ainda que possuam recursos para tanto.

Confira abaixo o mapa mental sintetizando os principais argumentos apresentados em relação à avaliação dos mecanismos automatizados:

Figura 2 - Principais argumentos positivos e negativos em relação à avaliação do uso de mecanismos automatizados para moderação de conteúdo



Sobre a necessidade de revisão humana de decisões automatizadas

Este tópico corresponde a uma das questões que derivaram do questionamento em torno do uso de mecanismos automatizados de moderação de conteúdo. As pessoas foram instigadas a se manifestar sobre como avaliavam a revisão humana em relação ao uso dessas ferramentas.

Nesse sentido, considerando também as pessoas que já haviam se manifestado de forma positiva no tópico acima, houve ampla manifestação **defendendo a possibilidade de revisão humana de decisões automatizadas (12x)**. Além de ser considerada **necessária (6x)**, ela foi apontada como importante para os casos em que há decisões capazes de afetar a esfera de direitos dos usuários.

Desse modo, a moderação automatizada poderia atuar como um mecanismo de priorização de conteúdo, **realizando uma primeira análise e identificando conteúdo potencialmente nocivo (3x)**, que seria então encaminhado para revisão humana. Tal argumento vai ao encontro dos entrevistados que defenderam que **as atividades de moderação por IA e por humanos fossem utilizadas de maneira complementar (2x)**.

Nesse contexto, também foi argumentado que a revisão humana deveria ser uma garantia aos usuários que **discordassem de eventual decisão automatizada sobre moderação e recorressem à plataforma para revisão (3x)**. Ainda, ela poderia atuar como uma segunda etapa para barrar o que eventualmente houvesse escapado ao filtro de mecanismos de IA. No ponto, dois entrevistados apontaram que houve uma perda em não garantir a obrigatoriedade da revisão humana por meio da Lei Geral de Proteção de Dados.

Outro desafio, apontado por Guilherme Araújo, do setor privado, foi no sentido de que haveria categorias de conteúdo muito dependentes de contexto, de modo que não existiria ainda uma tecnologia boa e eficiente o bastante para dispensar a necessidade de qualquer revisão humana. Por outro lado, Lucas Oliveira, do terceiro setor, destacou que é difícil assegurar a revisão humana de maneira escalável, considerando o grande volume de informações disponibilizado na internet a todo momento. Apesar disso, destacou que cabe não somente sociedade civil, mas às próprias plataformas, que possuem recursos para tanto, realizar pesquisas e desenvolver produtos nesse sentido, buscando

(...) como inserir um procedimento de revisão por humanos em decisões que foram tomadas única e exclusivamente por algoritmos de forma automatizada, pra que a gente tenha esse controle, essa segunda garantia de que de fato aquela decisão tá correta, e infelizmente hoje a gente ainda não tem essa garantia e acaba levando a distorções (...)

Entretanto, dentre os aspectos negativos mencionados em torno da revisão humana, estava a preocupação com o **bem-estar psicológico dos trabalhadores envolvidos na moderação de conteúdo (2x)** e a violação de seus direitos humanos, em virtude da exposição prolongada a conteúdos nocivos. Foi argumentado que se trataria de um subtrabalho, muitas vezes delegado a países em desenvolvimento, em virtude da possibilidade de as empresas pagarem menos a esses funcionários e não terem de lidar com as consequências provocadas por essa atividade. Ademais, haveria pouca informação sobre o funcionamento de times de moderação e seus mecanismos, incluindo aqueles de remediação a eventuais danos derivados do trabalho de revisão humana.

Para Lara Souza, do terceiro setor, é necessário pensar em como inserir esse debate na regulação da moderação de conteúdo também:

Como é que a gente - dentro dessa discussão, se a gente está debatendo a regulação da moderação de conteúdo, quais são as regras e tal -, **como que a gente também de alguma maneira pode prever certas garantias de direitos dessas pessoas que vão estar fazendo essa regulação, já que a gente já está vendo, em tão pouco tempo em que essa atividade existe, efeitos dentro da vida dessas pessoas, que eu não acho que sejam levianos?** Assim, não são irrelevantes, o que eu quero dizer. E que, assim, se a gente já vê um efeito tão grande dessa circulação de conteúdos nocivos dentro das plataformas, interferindo na nossa democracia, interferindo na nossa saúde mental, interferindo na polarização, na radicalização política - se a gente já vê isso dentro dos usuários que estão sob moderação, ou seja, os usuários que não estão, assim, não estão vendo esse tipo de conteúdo porque ele foi moderado, mas pra ele ter sido moderado talvez outra pessoa tenha visto esse conteúdo, né? E assim, o efeito que vai ter, psicológico, nesse grupo de pessoas, eu acho que isso tem que ser debatido, sim (grifo nosso).

Por fim, Theo Freitas, do terceiro setor, pontuou que somente garantir uma moderação por humanos não resolveria todos os problemas da própria moderação, alegando que a reivindicação por essa pauta passaria apenas por uma espécie de legitimação da decisão feita por sistemas automatizados.

4.2. Sobre o papel e atuação dos diferentes atores da governança da internet envolvidos na moderação de conteúdo

Neste tópico, estão as respostas dos entrevistados sobre a sua percepção em torno de qual seria o ator com maior responsabilidade sobre a gestão da moderação de conteúdo, mecanismos de fiscalização dessa atividade disponíveis para a sociedade civil, avaliação de instâncias privadas de revisão de decisões de moderação e da atuação do setor técnico-científico e terceiro setor. Contemplam-se as perguntas 4 a 7 do roteiro.

Sobre a responsabilidade pela gestão da moderação de conteúdo

Quando questionados sobre o agente³⁴ da Governança da Internet que detém maior responsabilidade diante da gestão da moderação de conteúdo, a grande maioria dos entrevistados se posicionou afirmando ser unicamente o setor privado, representado pelas plataformas e grandes conglomerados de tecnologia. Assim, as respostas se dividiram entre **apenas as plataformas (15x), plataformas e setor governamental (1x), apenas o setor governamental (1x) e responsabilidade conjunta entre todos os setores (1x)**.

Dessa forma, os dois principais argumentos utilizados para discorrer sobre o poder das plataformas diante da moderação de conteúdo foram: disponibilidade de **recursos (7x) e competência(12x)** para gerir tal atividade. Dessa forma, no que diz respeito ao primeiro argumento, tem-se que tais recursos abrangem uma série de elementos que levam tais empresas a uma posição vantajosa. Dentre eles: i) o poderio econômico, que possibilita a aquisição de tecnologia qualificada e, assim, capacidade para lidar com o grande fluxo de informações do ambiente virtual; ii) a posição de expertise conquistada em conjunto ao desenvolvimento tecnológico; e iii) a margem regulatória, que permite liberdade na criação de regras e diretrizes próprias - visto a ausência de diretrizes pontuais advindas do setor governamental.

Visto isso, a relação entre os recursos pertencente às plataformas e seus elementos é ilustrada por Igor Peixoto, membro do setor empresarial, da seguinte forma:

Tem a questão da **escala, acho que seria muito difícil transferir algumas dessas responsabilidades para outros entes**, inclusive, entes públicos, porque é uma quantidade imensa de conteúdo

34 incluindo setor privado, governamental, terceiro setor e comunidade técnico-científica.

sendo postado e essas plataformas, assim, principalmente as grandes plataformas - porque a gente também pode fazer uma diferenciação entre as grandes, as médias, as pequenas, as grandes plataformas, elas têm muito recurso também para conseguir lidar com essa questão da escala e também eu diria muito expertise, porque ao lidar com esse volume de conteúdo ao longo dos anos, **elas foram aprendendo a lidar com maiores eficiências para fazer com que o sistema como um todo funcione melhor** (grifo nosso).

Em complemento, o entrevistado também fez ponderações sobre como as plataformas, a partir de suas atividades e tempo de atuação, detêm uma quantidade de informações que seria inatingível pelos outros agentes. Logo, observa-se uma assimetria informacional e uma impossibilidade de transferência de *know-how* aos outros setores.

No que diz respeito ao segundo argumento apresentado pelos entrevistados - a competência das plataformas, - o ponto foi exemplificado a partir da noção de que a gestão sob a moderação de conteúdo seria uma atribuição natural de tais empresas.

Assim, visto a posição de hegemonia e flexibilidade regulatória ocupada por tais agentes, a condução das diretrizes de moderação seria uma responsabilidade inerente às suas atividades. Nesse contexto, Antonio Cavalcanti, membro do setor governamental, afirma que “a moderação de conteúdo é uma atividade **própria da camada de aplicações da rede**, ou seja, das empresas que disponibilizam conteúdo diretamente ou por meio de terceiros (grifo nosso).”

Ainda fazendo jus a esse pensamento, Theo Freitas, membro do terceiro setor, exemplifica como entende a gestão sob a moderação de conteúdo como uma responsabilidade inerente à atividade das plataformas da seguinte forma:

No limite, eu vejo essa atividade como uma atividade de responsabilidade das plataformas. É claro que, talvez o governo tenha uma função, o terceiro setor tenha uma função, a academia tenha uma função... Mas, nesse caso, **assim como, por exemplo, a neutralidade da rede, do Marco Civil, tem diretamente a ver com a atividade das companhias de telecomunicações, porque elas são as controladoras da estrutura, essa é uma atividade que tem totalmente a ver com a atuação das plataformas** (grifo nosso).

Em complemento, é possível afirmar que os argumentos quanto à disponibilidade de recursos apresentada pelas plataformas e sua competência diante da gestão da

moderação de conteúdo apresentam relação direta. Nesse sentido, Lucas Oliveira, membro do terceiro setor, mescla ambos os elementos ao discorrer sobre os motivos que, em sua opinião, tornam o setor privado o maior responsável sob a gestão da moderação de conteúdo ao dizer:

(...) Primeiro porque **elas criaram as próprias plataformas e elas conhecem a plataforma**, inclusive, de um **ponto de vista técnico**, então elas estão na melhor posição possível pra moderar aqueles serviços. Então quando, eventualmente, um novo problema surge, como foi o caso durante a pandemia de covid19, então você tem um novo fenômeno e a gente tem que enfrentar agora a desinformação sobre vacina, desinformação sobre lockdown e sobre o vírus em si também. **As plataformas precisam atualizar os seus protocolos de moderação, os seus próprios padrões e regras de uso daquele determinado espaço digital, e elas que têm que fazer isso, porque elas estão na melhor posição possível para fazer isso e acabam recebendo também a maior parte da responsabilidade pela moderação justamente porque estão ali na linha de frente, no front da batalha contra conteúdos nocivos** (grifo nosso).

No que diz respeito à questão da previsão legal e sua relação com a competência das plataformas, a relação desenvolvida foi em relação à ausência de diretrizes expressas quanto à atuação do setor privado diante da moderação de conteúdo e, logo, o surgimento de uma grande margem de atuação. No entanto, Fernanda Rezende, membra do terceiro setor, defendeu que, apesar da ausência de uma lei específica para a definição e aplicação de regras no campo da moderação de conteúdo, existem marcos legais em no ordenamento jurídico brasileiro que atribuem às plataformas a responsabilidade sobre a gestão de tal atividade, mesmo que de maneira indireta. Seria então o caso do Código de Defesa do Consumidor, o qual caracteriza as empresas como prestadoras de serviços e, logo, as submete ao seu regime de responsabilidade. Fernanda comentou também sobre o regime assegurado pelo Marco Civil da Internet e pela Lei Geral de Proteção de Dados, a qual estaria intrinsecamente ligada à moderação de conteúdo em vista às ações de coleta e tratamento de dados.

Ademais, a posição de hegemonia das plataformas e o poderio na gestão da moderação de conteúdo frente aos demais agentes que compõem a Governança da Internet também foram comentados sob um viés negativo por alguns entrevistados. Nesse sentido, Caio Fernandes, membro do setor governamental, pondera que

às vezes as grandes plataformas são uma **distorção à internet original, porque concentraram poder, concentraram usuários**

e porque nós somos preguiçosos de procurar as coisas na nossa... comodistas, né, então, a gente acaba aceitando esse tipo de coisa, mas isso não é bom, não é bom pelo enorme poder que eles... ameilham nesse processo (grifo nosso).

Por outro lado, a questão do protagonismo das plataformas também foi comentado sob um ponto de vista distinto. Nesse sentido, Guilherme Araújo, do setor privado, pontuou que

Há um debate bastante moderno sobre se essa responsabilidade deveria ser única das plataformas ou não, mas o fato é que da forma como isso funciona hoje, **a responsabilidade recai única e exclusivamente nas plataformas. Com o que isso tem de bom e com o que isso tem de ruim** (grifo nosso).

Em vista à grande concentração de poder das plataformas, alguns entrevistados se posicionaram quanto à **necessidade de haver uma atuação mais impositiva por parte do setor governamental (7x)**. Nesse sentido, Sofia Pires, membra do terceiro setor, destaca a necessidade de incidência e regulação por parte do governo em prol de mitigar o caráter negligente e abusivo das ações do setor privado. Em complemento, Augusto Cruz, membro do setor governamental, argumenta que as próprias plataformas devem atuar na moderação; no entanto, é necessário que os Estados Nacionais regulamentem de alguma forma, traçando parâmetros de máximo e mínimo.

Quanto à posição de que as plataformas dividem a responsabilidade sobre a gestão da moderação de conteúdo com o setor governamental, Luiz Porto, membro da comunidade técnico-científica, apresentou reflexões sobre os distintos poderes e atuações que os setores privado e governamental apresentam enquanto tomadores de decisão. Nesse sentido, pontua que o Estado se destaca por apresentar um poder decisório único, atribuído pela jurisdição. A partir disso, goza de mecanismos legais que as plataformas, enquanto representantes do setor privado, carecem. Em contraponto, os grandes conglomerados possuem um poder corporativo de caráter transnacional, o qual garante o poder de elaboração e aplicação das políticas internas; apesar deste estar submetido ao poder governamental.

Já no que diz respeito à posição de que o setor governamental seria o principal agente diante da gestão da moderação de conteúdo, Maria Teixeira argumenta:

Olha, eu acho que **os Poderes da República, no nosso caso, compartilham essa maior responsabilidade**. Até porque eu não acredito que o Conselho de Autorregulação das plataformas funcionaria (...) eu acho que os Poderes da República

precisam dividir essa responsabilidade (...) Precisa haver um **desenvolvimento de projetos de lei pra gente garantir primeiro a transparência dessa moderação de conteúdo, segundo a padronização dessa moderação de conteúdo e ter sempre a garantia de que ela vai acontecer de acordo com a legislação** (...) Eu acho que os três atores da república são os três poderes mais importantes, isto porque eles têm o poder do **constrangimento legal**, digamos assim, sobre as plataformas (grifo nosso).

Por fim, tem-se a posição apresentada por Bárbara Silveira, representante do setor empresarial, que argumenta em prol de uma responsabilidade conjunta de todos os setores diante da gestão da moderação de conteúdo. Bárbara afirma que, no que diz respeito ao papel das plataformas, cabe a elas buscar por maneira inovadoras de lidar com as questões que possam superar a abordagem binária da remoção ou manutenção de conteúdo, devendo elas também operar de maneira direta na mudança de comportamentos ou mesmo na prevenção por meio de incentivos ou desincentivos.

Já no que diz respeito à atuação por parte do setor governamental, diz que “os governos podem e devem priorizar políticas, parcerias, investimentos domésticos e no exterior que apoiem a alfabetização digital e defendam a Internet Aberta”. Por fim, argumenta que a sociedade também possui um papel fundamental quanto à educação midiática das pessoas, “uma vez que a desinformação apenas funciona se houver uma audiência que engaje com ela”. Em adição, caberia também à sociedade a fiscalização diante do progresso dos esforços dos governos e das empresas, além do oferecimento de contribuições para melhoria de tais esforços e cobrança para que tais estejam alinhados aos padrões internacionais de direitos humanos.

Sobre possíveis meios de fiscalização do processo de moderação de conteúdo

O presente tópico visa a analisar as falas dos entrevistados quando questionados sobre a existência de mecanismos para garantir a fiscalização, por parte da sociedade, diante de práticas de moderação de conteúdo. Ou seja: Existiriam maneiras e recursos que permitem à sociedade acompanhar como as práticas de moderação de conteúdo estão sendo executadas?

A grande maioria dos entrevistados se apresentou de forma cética à eficiência ou mesmo à existência de tais mecanismos. Assim, as classificações observadas foram: **inexistentes (7x); existentes, mas insuficientes(6x); existentes, mas com ressalvas em sua eficiência (3x) e suficientes (1x)**. Por fim, um entrevistado afirmou apenas que tais são necessários. Dentre os exemplos de meios utilizados para a fiscalização das práticas de moderação de conteúdo, **os relatórios de transparência foram os mais citados (6x)**.

Dessa forma, os principais argumentos apresentados para defender a inexistência de tais mecanismos foram: a ausência de diretrizes regulatórias que obriguem as plataformas a desenvolvê-los e a discricionariedade e liberdade desses agentes no momento de criá-los, visto que apresentam caráter opaco e seletivo - não podendo assim serem considerados como meios de fiscalização por parte da sociedade.

Nesse sentido, Ana Carvalho, membra do terceiro setor, afirma:

Então, eu não sei se existe, eu acho que **não existe mecanismo que garanta com absoluta certeza que a sociedade civil vai poder fiscalizar o que acontece ali**. Porque, ao mesmo tempo, eles só vão divulgar o que interessa, o que eles querem que a gente saiba. Então, ao fim e ao cabo, o poder sobre essa atividade de moderação de conteúdo está com eles, com as empresas, e muito pouco com a gente nessa capacidade fiscalizatória assim. **Por mais que a gente faça essa imersão nos relatórios de transparência, termos de serviço, nas políticas de privacidade... acho que não existe um mecanismo, até porque as atividades de moderação são muito não explicadas, ou não explicáveis, pelo menos o que a gente vê na prática**. Não tem muito algo tipo “você infringiu tal norma do termo de serviço e, por isso, a sua conta está sendo suspensa por uma semana”. Isso não existe, não acontece. Então a gente fica muito à mercê das empresas (grifo nosso).

Em complemento, Theo Freitas, também membro do terceiro setor, discorre sobre a ausência de diretrizes expressas quanto à criação de tais mecanismos de fiscalização da seguinte forma:

Melhor seria se a gente tivesse uma regra mais clara dizendo o que pode e o que não pode nessa atividade e, ainda melhor seria se a gente tivesse um **órgão ou uma atividade independente que não pudesse entrar na discussão de conteúdo**, porque eu acho isso muito perigoso, mas que pudesse entrar nessa análise mais sistêmica, digamos assim, dessa atividade, nos moldes do DSA, possivelmente, né, na legislação europeia, é possível (grifo nosso).

Nesse sentido, Samuel Cardoso, membro do setor técnico-científico, comenta que não apenas no Brasil, mas em todo o mundo, há um regime jurídico tênue, leve e permissivo em relação à exigência de transparência e de prestação de contas. Dessa maneira, não haveria como a sociedade saber se as práticas adequadas no que tange à moderação de conteúdo estão de fato sendo executadas.

Ainda no que diz respeito à questão regulatória, Stephany das Neves, membra da comunidade técnico-científica, pontua que “ainda estamos em um estágio anterior a isso” no que diz respeito à discussão e análise sobre o tema. Nesse contexto, pontua:

(...) eu acho que quando o Marco Civil trouxe a regra da isenção de responsabilidade a partir da **decisão judicial, mas sem ter trazido a regra do bom samaritano, a gente acabou não trazendo essa cultura de que as plataformas por um lado precisam moderar e podem moderar como se fosse uma faculdade de dever**. Mais do que uma faculdade de dever, é algo quase inerente do próprio serviço digital. A moderação faz parte desse serviço. Eu acho que isso não está claro em nossa cultura jurídica, eu acho que **o fato do Marco Civil ter silenciado sobre isso também não ajudou**, mas eu acho que não é só o Marco Civil, a gente também talvez como comunidade acadêmica, científica, técnica, todos nós. Esse tema chegou agora, mas nós temos quase dez anos de Marco Civil, 2014, faremos daqui um ano e meio, em 2024. Então eu acho que **a gente não discutiu esse tema suficientemente e com isso eu acho que de fato a gente sequer... é uma atividade que é aceita hoje como algo completamente normal** (grifo nosso).

No que tange aos entrevistados que entendem os mecanismos de fiscalização como existentes, mas insuficientes, o principal argumento para defender tal posição foi quanto ao caráter opaco das manifestações, políticas e relatórios de transparência divulgados pelas plataformas, de maneira que não há uma transparência real. Assim, várias críticas foram direcionadas aos relatórios de transparência, especificamente, a exemplo dos comentários realizados por Lucas Oliveira, membro do terceiro setor:

Olha, existem alguns mecanismos, tá? Só que a gente ainda depende muito da boa vontade das próprias plataformas e isso é um problema. O que eu quero dizer com isso: **são as plataformas que eventualmente vão publicar um relatório de transparência, onde você vai ter informações sobre as práticas de moderação de conteúdo de um ponto de vista sistêmico e não caso a caso** (...) E muito do que a gente tem são as próprias plataformas que estão disponibilizando isso de forma voluntária, então a gente ainda precisa avançar muito em termos de transparência da moderação de conteúdo (grifo nosso).

Outro ponto apresentado por Lucas Oliveira, por sua vez, foi quanto ao papel do Estado na função de exigir informações “completas e certas” por parte das plataformas, visto que não são todas que interessam ao público. A partir disso, seria possível aos usuários alcançarem uma postura mais proativa e de cobrança frente ao setor privado.

Antonio Cavalcanti, membro do setor governamental, também opina em prol da insuficiência dos mecanismos oferecidos pelas plataformas ao dizer que:

(...) embora algumas empresas tenham seus sistemas internos de apuração, inclusive aceitando reclamações e recursos de suas decisões, esses procedimentos **não são uniformes nem transparentes para grande parte dos usuários**. Na maioria das vezes, a resposta a essas apelações é imprecisa, vaga, insatisfatória ou mesmo inexistente (grifo nosso).

Quanto aos entrevistados que argumentaram em prol da existência de mecanismos de fiscalização, mas com ressalvas em sua suficiência, Guilherme Araújo, membro do setor privado, defende que tais meios ainda estão em processo de desenvolvimento e, embora não tenhamos atingido o estado da arte (sic), o contexto atual seria significativamente melhor do que aquele de 4 ou 5 anos atrás. Por outro lado, Vicente Moura, do setor governamental, afirma que a suficiência dos mecanismos de fiscalização depende de uma “ação combinada”. Nesse sentido, diz:

Eu acho que um primeiro passo são relatórios de transparência acerca do número de retiradas de conteúdo, de moderações discriminadas por tipo, mas eventualmente também a própria procedimentalização desses processos dentro da própria plataforma, então eventualmente **dar a chance ao usuário para justificar por que aquele post deve ou não deve ser moderado, tanto do lado de quem denuncia quanto do lado de quem tem o seu post moderado. Acho que seriam esses instrumentos** (grifo nosso).

Por fim, Bárbara Silveira, membra do setor empresarial, afirma que os mecanismos de fiscalização existem, mas podem melhorar. Nesse sentido, pontua que um dos mecanismos de fiscalização é a transparência, a qual garante o accountability por parte das empresas e governos e, logo, a responsabilização. Ainda, adiciona que uma área passível de melhora é quanto à garantia de que “as leis que regem as informações ofereçam flexibilidade adequada para divulgações valiosas, por exemplo, o fornecimento de dados a acadêmicos e pesquisadores.”

Por fim, os argumentos apresentados em prol da existência e eficiência dos meios de fiscalização diante das práticas de moderação de conteúdo focam nos mecanismos de natureza extrajudicial e judicial no âmbito da proteção dos direitos de usuários da internet. Nesse sentido, Luiz Porto, membro da comunidade técnico-científica, afirma:

(...) e aí olhando especificamente se nós observamos a possibilidade de que haja **tutela coletiva**. Eu tô pensando na tutela coletiva, porque é pela tutela coletiva que a sociedade civil busca as formas de proteção. Seja, individualmente, por meio das **ações civis públicas, por intervenção de terceiros, seja pela participação em ações civis públicas com o Ministério público federal estadual**, a depender das competências que são trazidas em consideração (...) Ou administrativo judicial, por força da jurisdição do estado, ou moral, quando há, por exemplo, **boicote coletivo ao uso de determinados serviços ou a determinados produtos, pensando aí na condição de usuário, de consumidor, de cliente, de cidadão**, é... então, seria uma outra forma dentro de uma **consequência fiscalizatória**, mas hoje eu acho que é importante perceber que há mecanismos mais ou menos fortes no campo de tutela coletiva, e aí depende, porque muitas vezes não é interesse, por exemplo do Ministério Público acompanhar essas questões (grifo nosso).

Observa-se assim que Luiz Porto interpreta os meios de fiscalização sob uma diferente perspectiva daquela levada em conta pelos outros entrevistados.

Sobre instâncias privadas de revisão de moderação de conteúdo

No que diz respeito às instâncias privadas de revisão de moderação de conteúdo, foram considerados os órgãos de iniciativa privada que se propõem a atuar como um revisor externo diante das decisões tomadas pelas plataformas, principalmente. Visto isto, ao questionar os entrevistados sobre seus posicionamentos quanto a tais instâncias, o Oversight Board (OB), órgão criado pela Meta, foi utilizado como exemplo para ilustrar a pergunta. Nesse sentido, a grande maioria dos entrevistados **adotou o OB como referência principal ao discorrer sobre o tema (14x)**. Ademais, cabe mencionar que **uma parte dos entrevistados realizaram ressalvas no início de suas falas (5x), pontuando:** ou que não acompanha de perto as decisões do OB; ou que tem em mente que poucas decisões foram tomadas e, assim, ainda não haveria muito a concluir nesse momento; ou que não conhece muito sobre tal órgão.

Dessa forma, tem-se que, quando questionados sobre sua posição diante das instâncias privadas de revisão de moderação de conteúdo, grande parte dos entrevistados afirmou que tais apresentam uma **atuação predominantemente positiva (12x)**, contra uma minoria que defendeu a **atuação predominantemente negativa (5x)** e um entrevistado que não tomou posição. De maneira complementar, foi também questionado se a atuação de tais instâncias garantiria uma **maior legitimidade frente às atividades de moderação de conteúdo, ao que parte dos entrevistados afirmou que sim (6x)**, alguns se posicionaram de **maneira contrária (3x)**, uma parcela opinou quanto à uma **legitimidade não determinada (4x)** e a parcela final **não comentou sobre (5x)**.

Partindo à análise das respostas, no que diz respeito aos principais argumentos apresentados para defender a atuação positiva das instâncias privadas de revisão de moderação de conteúdo, tais foram: a iniciativa em prol da desconcentração de poderes no momento de tomada de decisão e a garantia de maior transparência nos procedimentos de moderação de conteúdo.

Nesse sentido, Theo Freitas, membro do setor privado, argumenta:

Eu avalio positivamente. Eu acho que foi um passo... eu acho que foi um passo importante. É um **certo experimentalismo para lidar com isso**. Sem dúvida, há uma discussão sobre o quanto você está terceirizando responsabilidade, mas eu não acho que isso seja necessariamente ruim, porque no limite do limite, você não está terceirizando responsabilidade sobre a atividade industrial, você tá **meio que tentando ajuda nos casos mais difíceis**, ganha transparência... e é um **ganho de transparência (...)** É um ganho de transparência na justificação das decisões, não nas decisões em si. Nesse sentido, eu acho um ganho (grifo nosso).

Em complemento, Lucas Oliveira, membro do terceiro setor, argumenta em prol da atuação do Oversight Board, ressaltando a legitimidade por ele assegurada às decisões do Meta ao descentralizar o procedimento de tomada de decisão:

Eu acho que, assim, é uma iniciativa pensada justamente para tentar trazer mais legitimidade ao processo de moderação de conteúdo. O que você tinha antes era basicamente uma concentração de poderes, então, assim, altos executivos na empresa, eles tinham os 3 poderes em 1: eles eram legisladores, julgadores e administradores das plataformas ao mesmo tempo, realizavam essas 3 funções. (...) E o Oversight Board do Facebook é uma dessas iniciativas de separação de poderes, porque você cria um braço de revisão de decisões de moderação de conteúdo em última instância que é independente da empresa, então é como se você tirasse um poder que antes estava concentrado nos altos executivos da empresa e agora você dá pra um corpo, um comitê independente de especialistas que vai fazer esse julgamento em última instância e **a empresa é obrigada a respeitar a decisão final**. Isso dá mais legitimidade ao processo de moderação de conteúdo justamente porque você começa a descentralizar um pouco essa tomada de decisão (grifo nosso).

Stephany das Neves, membra da comunidade técnico-científica, apresenta argumentos semelhantes em defesa da garantia de maior legitimidade assegurada pelas instâncias privadas de revisão. Nesse sentido, diz:

É natural que, para que a decisão seja mais legítima, que ela não seja tomada só por alguém da empresa, como se fosse o CEO ou o diretor, diretor jurídico ou algo assim. Mas que de fato tenha mais legitimidade, especialmente a **legitimidade dada pelo próprio procedimento.** Talvez a legitimidade dada pela **representatividade desse conselho,** pelo fato dela ser externa e o procedimento em si, eu acho que ele pode trazer a legitimidade. Então certamente impacta positivamente (grifo nosso).

No que tange à garantia de maior transparência a partir das ações do Oversight Board, Igor Peixoto, membro do setor privado, argumenta que, embora as recomendações publicadas pelo órgão não sejam vinculantes, tais diretrizes conferem maior visibilidade às atividades de moderação de conteúdo. Assim, caso a Meta deixe de aplicá-las, não haverá determinação de sanção por parte do órgão; no entanto, pode haver pressão por parte dos grupos afetados.

No que diz respeito àqueles entrevistados que consideram a atuação de tais instâncias predominantemente negativa, os argumentos mais recorrentes foram quanto à: **limitações na autonomia ou atuação(12x), pouca transparência na tomada de decisões e criação de diretrizes(3x), baixo impacto de suas atividades(5x) e alto custo para a implementação e funcionamento(2x).** Tais argumentos, por sua vez, também foram apresentados por alguns entrevistados que se posicionaram de maneira favorável à atuação de tais instâncias, tendo sido pontos a serem melhorados ou simplesmente críticas realizadas.

No que tange às limitações de autonomia e atuação, os entrevistados discorreram sobre a criação das instâncias privadas serem iniciativas das próprias plataformas, o que, por sua vez, impediria a tomada de decisões de fato imparciais. Nesse sentido, também pontuaram sobre os próprios funcionários serem escolhidos pelas plataformas e, recorrentemente, tais órgãos, apesar de tidos como externos, serem vinculados às empresas em questões sócio-políticas e reputacionais (impedindo uma atuação independente).

Nessa seara, é possível observar que, mesmo os entrevistados que se posicionam em prol da atuação dessas instâncias, tais críticas estão presentes. Isso pode ser ilustrado pela fala de Fernanda Rezende, membra do terceiro setor, ao dizer que, apesar de positivas, elas são agentes insuficientes pois “estão estritamente **relacionadas e organizadas e dirigidas pela própria empresa**” (grifo nosso).

Da mesma forma, todos os entrevistados que argumentaram em prol da atuação negativa das instâncias de revisão de tomada de decisão também acreditam que tais órgãos não garantem uma maior legitimidade no processo de tomada de decisões no campo da moderação de conteúdo. Nesse sentido, a fala de Samuel Cardoso, membro do terceiro setor, ilustra bem tais questões:

Eu acho que a afetação da legitimidade, assim, **tem um risco de esses órgãos funcionarem apenas para limpeza da imagem, num sentido de a empresa fazer uma coisa de fachada e não de fato ser implementado**, e talvez também **pelo status de compor um órgão, alguns integrantes, em prestígio à empresa, acabem não impondo as suas decisões**. Eu acho que existe uma chance eventual de esse resultado ser positivo, mas eu acho que não tem exatamente garantias, então eu acho que por mais que houvesse espaços para uma decisão mais eficiente é muito possível que ela não aconteça assim. Eu acho que eventualmente entre as empresas, até quando abrem esses espaços, elas podem estar dando legitimidade de esperar que de fato vendem algum posicionamento mais honesto assim, mais sincero, né, **mas eu acho que a estrutura e a dinâmica de formação desses grupos acabam limitando a legitimidade que eles têm pra poder...** propor, sugerir medidas mais drásticas ou proporcionais à gravidade da situação (grifo nosso).

Os comentários de Samuel, por sua vez, encontram elementos em comum na fala de Antonio Cavalcanti, membro do setor governamental, ao dizer que:

Essas instâncias privadas, não obstante, sejam uma demonstração de boa-fé além de desempenharem papel importante na percepção de que as plataformas estão ‘fazendo alguma coisa’, por assim dizer, **não gozam de transparência e legitimidade suficientes para ganhar a confiança plena dos usuários** (grifo nosso).

Uma questão recorrente comentada tanto pelos entrevistados que consideram a atuação de tais instâncias positiva, como os que a consideram negativa, foi quanto ao baixo impacto alcançado pelas suas atividades. Assim, visto o volume da demanda que as grandes plataformas possuem, órgãos externos não seriam capazes de “resolver o problema como um todo”, como afirma Ana Carvalho, membra do terceiro setor. Além disso, a entrevistada pondera que, em relação ao Oversight Board, apesar do órgão ser uma boa iniciativa, apenas os casos paradigmáticos seriam por ele considerados, ignorando um grande volume de casos corriqueiros e mantendo medidas controversas e de possível ofensa à liberdade de expressão.

Nesse mesmo sentido, Guilherme Araújo, membro do setor privado, opina:

(...) isso é melhor do que nada, a gente... mas eu acho que **a gente ainda não tem condições de avaliar a efetividade desse órgão, seja porque ele julgou poucas coisas, seja porque de fato a gente tem pouca produção**, ah... a respeito de como é que ele tá funcionando, não no varejo, a gente sabe que algumas medidas foram revertidas, outras mantidas, mas no atacado, eu não creio que a gente tenha ainda condições de fazer uma avaliação sobre esse modelo (grifo nosso).

Nesse contexto, cabe mencionar dois entrevistados que comentaram sobre essa característica dos órgãos argumentando em um sentido distinto. Na visão deles, haveria um mal entendimento das pessoas sobre a proposta apresentada, de maneira que o objetivo não seria reanalisar todos os casos em larga escala, mas sim traçar grandes diretrizes justamente a partir da análise de casos mais difíceis ou paradigmáticos. Ao defender esse ponto de vista, Theo Freitas, afirma que há “**uma dissonância sobre o que as pessoas querem de um órgão como esse [Oversight Board] e o que que esses órgãos realmente querem oferecer**” - que, em sua visão, seria mais clareza e transparência à plataforma.

Em complemento, Igor Peixoto, membro do setor privado, comenta sobre o impacto do Oversight Board e suas críticas buscando explicar sobre a questão da (não) vinculação no que diz respeito às recomendações e decisões publicadas:

E acho que também outro aspecto que suscitava alguma desconfiança em relação ao real impacto do Board é o fato de que as recomendações, elas não são vinculantes e isso significa que **a empresa não é compelida a aceitar essas recomendações. As decisões de remover o conteúdo ou não, essas, sim, são vinculantes, mas as recomendações não**. Só que, apesar disso, a empresa, a Meta, ela precisa revelar publicamente quais recomendações ela vai implementar e também quais ela não vai implementar e os motivos por trás da decisão de implementar uma decisão ou não implementar. E eu entendo que só já nessa troca, assim, **o montante de informação divulgada sobre o funcionamento interno da atividade de moderação como, isso é algo positivo, do ponto de vista de transparência, que eu diria que é meio único dessa relação do Board com a Meta, até esse momento, não tem algo parecido nas outras plataformas** (grifo nosso).

Por fim, uma crítica ao Oversight Board apareceu uma única vez, mas apresenta relação direta com outras questões pontuadas. Trata-se de como acontece o engajamento entre o OB e atores externos, questão comentada por Lucas Oliveira. Para o entrevistado, o órgão tem enfrentado dificuldades em “conversar” com os comentários públicos realizados pelo público interessado (que pode ser uma instituição ou qualquer indivíduo em sua capacidade pessoal). Comenta ainda que este problema não é exclusivo do Board, mas recorrente em tribunais superiores ao redor do mundo, a exemplo do Supremo Tribunal Federal, ao não levar em consideração as manifestações realizadas por *amicus curiae* em audiências públicas. A superação desta questão, de acordo com Lucas, conferiria maior legitimidade ao órgão, ao passo que as decisões passariam a ser construídas de forma coletiva.

Sobre o papel do Poder Judiciário enquanto possível instância de revisão judicial de decisões envolvendo moderação de conteúdo

Neste tópico, estão as percepções dos entrevistados sobre o papel do Poder Judiciário como uma possível instância de revisão judicial para decisões de moderação de conteúdo. A realização desse questionamento se deu pela possibilidade de os usuários buscarem o acesso à justiça para reverter eventual remoção de conteúdo ou suspensão de conta, por exemplo.

Nesse sentido, o Marco Civil da Internet foi citado como um dos responsáveis por tornar o Poder Judiciário como **parte essencial quando se fala em moderação de conteúdo (3x)**, em virtude do previsto no art. 19, que estabelece a responsabilização dos intermediários pelo conteúdo publicado por terceiros somente no caso de descumprimento de ordem judicial, salvo exceções legais específicas. Para Sofia Pires, do Terceiro Setor, a atuação para esses casos seria justa e necessária, tendo esse Poder como “terceira parte, capacitada e proporcional o bastante para poder decidir sobre esses casos”.

Guilherme Araújo, do setor privado, destacou ser comum a destinação de ordens judiciais para plataformas para reversão de decisões envolvendo moderação, sendo que o maior número de casos de disputa sobre o tema seria no Brasil. Tal afirmação vai ao encontro de argumentos levantados de que o Judiciário atua na **solução de conflito ou restrição de direitos (2x)**, análise de fatos com repercussões jurídicas, como a moderação de conteúdo, na **reparação de danos eventualmente causados por essa atividade (2x) e para correção de ilegalidades (2x)**. A esse respeito, ressalta-se o posicionamento de Stephany das Neves, do setor técnico-científico:

Então eu acho que ele tem um papel, sim, **que é identificar em determinados tipos de excesso na moderação, seja porque ele tirou conteúdo demais ou seja porque não tirou conteúdo**

indevido que deveria ter tirado, eu acho que o Judiciário tem um papel, sim, de realmente corrigir ilegalidades (grifo nosso).

Para Lara Souza, do terceiro setor, essa instituição também teria um papel fundamental, em especial, em período eleitoral, na moderação de postagens de políticos, contribuindo para uma maior legitimidade da própria moderação. Ainda, outro argumento apresentado nas entrevistas foi de que não haveria como escapar à atuação do Judiciário, em razão de garantias constitucionais de acesso à justiça. No ponto, Emanuella Nogueira, do Terceiro Setor, defendeu a impossibilidade de limitação desse acesso, como, por exemplo, condicionando à necessidade de um pedido prévio de revisão à plataforma. Entretanto, ela destaca que, em um cenário padrão, com situações mais regulares, a moderação deveria poder funcionar sem que fosse necessário levar provocações ao Judiciário, que poderia atuar somente nos casos mais graves.

Essa preocupação ganha ainda mais relevo diante do reconhecimento de que encaminhar um grande número de demandas de moderação de conteúdo para revisão judicial **poderia agravar a sobrecarga que tribunais já vêm enfrentando no país (3x)**. Além disso, dado o volume de informações disponíveis na internet, assim como a escala de moderação, seria inviável levar todas as disputas para as cortes, considerando também que fugiria à sua competência atuar enquanto uma instância de revisão dessa atividade. O comentário de Lucas Oliveira, do Terceiro Setor, ilustra bem essa percepção:

(...) **o Poder Judiciário também não pode atuar como uma instância de revisão do processo de moderação de conteúdo em si, aí foge à competência dele**. Ele não pode dar pitacos sobre como funciona a moderação de conteúdo, o que que pode melhorar e o que que não pode. Então, assim, funciona pra alguns casos específicos, quando você tem essa colisão de direitos, pra evitar, inclusive, maiores danos ou pra reparar danos que já aconteceram, **mas pra melhorar o processo de moderação de conteúdo enquanto sistema, aí o Poder Judiciário não pode fazer muita coisa** (grifo nosso).

Outro posicionamento negativo foi na direção de que o Poder Judiciário é uma instituição mais lenta, que necessita de provocação para poder atuar em casos específicos, razão pela qual apresenta um papel limitado e não pode ser a única alternativa. Ademais, para dar uma resposta melhor a esse tipo de litígio, **seria necessária maior dinamicidade (x)**, como por meio da criação de juizados especiais ou varas especializadas. Esta última alternativa poderia auxiliar também a mitigar outro ponto levantado pelos entrevistados, que foi o **risco de concessão de decisões judiciais inadequadas ou pouco técnicas (2x)**.

Para Samuel Cardoso, do setor técnico-científico, é preciso que decisões sobre moderação tenham “um outro grau de cautela por conta das consequências possíveis da permissão ou da negação, da concessão ou da negação de determinado pedido, considerando, inclusive, a possibilidade de o juiz realizar uma concessão parcial”. Da mesma forma, foi apontado que seria necessário à magistratura reconhecer o valor da atividade de moderação de conteúdo em suas decisões e compreender que nem toda moderação precisa ser evitada.

Ainda, foi mencionado que o Poder Judiciário, apesar de ter um papel importante, seria **mais um ator na equação da moderação de conteúdo, e não necessariamente o principal (2x)**. Nessa senda, Augusto da Cruz, do setor governamental, argumentou não entender como um pré-requisito para a remoção de conteúdo online a necessidade de recorrer ao Poder Judiciário, mas que desde que bem fundamentado na legislação, seria mais interessante garantir direitos ao usuário ou a reparação, pela via judicial, quando for o caso.

Outrossim, buscar o acesso à justiça envolveria alguns custos aos usuários, como **recursos financeiros e tempo para ingressar com uma ação judicial (2x)**. A falta de uma regra específica e debate prévio também prejudicariam a atuação do Judiciário, uma vez que aumentaria o risco das decisões não surtirem os efeitos esperados.

Para Caio Fernandes, do setor governamental, é necessário, ainda, que a criação de regras se dê sobre um “conceito de justiça local e cultura local do que pode e o que não pode”. No ponto, Samuel Cardoso defende a importância do processo legislativo para uma melhor atuação do poder judicial:

(...) eu acho que ele vai ser **tão mais eficiente quanto melhor tiver sido o processo legislativo de definição das regras**, né, que de antemão tenham empreendido o debate de qual que é a principiologia, né, o que que é o valor que vai ser defendido, né, pra cada [áudio inaudível] de moderação, pra cada situação, o que que tem ser coibido, do que que tem que ser valorizado, porque aí, mesmo que a decisão seja equivocada, ela vai ter que abordar esse aspecto, vai ter que dizer exatamente por que que ela não está aplicando aquela designação, qual o fator que ela está entendendo que é distintivo e que autoriza não ir em um sentido que está determinado pela legislação (grifo nosso).

Além disso, outros argumentos levantados apresentaram mais opções que contribuiriam para o papel desempenhado por este ator: políticas mais claras e maior transparência sobre casos de moderação de conteúdo; insistir em mecanismos previstos no Projeto de Lei 2.630/2020 (“PL das Fake News”), como para usuários disputarem decisões específicas de moderação em que se sentiram lesados; e investir em instâncias revisoras independentes, como o Oversight Board.

Entretanto, Ana Carvalho, do terceiro setor, pontuou ser importante chamar o Judiciário para resolver problemas envolvendo moderação, uma vez que não seria possível “só aceitar que a empresa toma uma decisão mais favorável a ela, infringindo direitos fundamentais como liberdade de expressão e privacidade”. Para Maria Teixeira, a importância da via judicial se justifica pelo constrangimento legal, capaz de fazer as plataformas agirem quando necessário. Desse modo, o seu papel seria atuar em conjunto com o Poder Legislativo, para fazer valer projetos de lei aprovados e atualizar códigos já existentes.

Sobre a participação do terceiro setor e academia nos processos regulatórios de conteúdo online

O presente tópico se propõe a reunir a opinião dos entrevistados a respeito da atuação e participação do terceiro setor e academia (ou comunidade técnico-científica) nos processos regulatórios de conteúdo online. Assim, foram contempladas as atividades realizadas por ambos os grupos no que diz respeito ao tema da moderação de conteúdo por meio de diferentes ações, a exemplo de incidência política, produção acadêmica, divulgação científica, participação em marcos regulatórios, dentre outros.

Diante disso, tem-se que a quase totalidade dos entrevistados opinou quanto à **participação predominantemente positiva do terceiro setor e academia (16x)** - referindo-se aos ambos setores em conjunto, um entrevistado opinou quanto à participação predominantemente negativa de ambos e um entrevistado argumentou em prol da atuação positiva do terceiro setor e negativa da academia.

No que diz respeito aos comentários positivos citados pelos entrevistados, foram mencionados: **qualificação do debate público (10x)**, participação significativa na **construção legislativa de marcos regulatórios (6x)** - como o Marco Civil da Internet, Lei Geral de Proteção de Dados e Projeto de Lei nº 2630/20, execução de **pressão para com plataformas e governo quanto ao cumprimento de diretrizes (4x)** e postura de **intermediação entre os setores (1x)**.

Assim, no sentido da qualificação do debate público e realizar intermediação entre os setores da Governança da Internet, Igor Peixoto, membro do setor privado, pontua:

Do ponto de vista do discurso regulatório, no sentido mais público da palavra, continua também sendo superimportante, porque **qualifica o debate público**. A gente sabe que muitas vezes **falta conhecimento técnico por parte dos parlamentares, querendo tirar soluções da cartola, coisas que a gente sabe que na prática não funciona e aí por isso, eu acho que essa interação com a sociedade civil e com a academia ajuda demais a qualificar**

o debate, a tirar algumas dessas percepções de que alguns problemas são fáceis demais de resolver (grifo nosso).

Nesse contexto, muitos entrevistados comentaram sobre a atuação do terceiro setor, principalmente, em processos legislativos. Nesse sentido, Fernanda Rezende, membra do terceiro setor, comenta sobre o processo de desenvolvimento do Marco Civil da Internet, da Lei Geral de Proteção de Dados e do surgimento da Coalizão de Direitos na Rede³⁵:

(...) a participação da sociedade civil foi bastante grande, **houve uma mobilização muito grande de diversas entidades da sociedade civil, e que deu origem inclusive à Coalizão**. A Coalizão surgiu desse coletivo que se formou nos debates do Marco Civil da Internet. E a contribuição da sociedade civil foi determinante para que, por exemplo, o Marco Civil da Internet não viesse com viés criminal, voltado para questões penais, com viés punitivista. Quem transformou o projeto inicial do Marco Civil da Internet - de uma lei que regulasse os direitos na internet - que inicialmente era um projeto com viés muito penal, **fomos nós, da sociedade civil, sociedade civil organizada que contribuiu de uma forma determinante pra que a gente chegasse no Marco Civil da Internet, um Marco que estabelecesse princípios, direitos gerais, para utilização da internet**. E a gente vem evoluindo nesse sentido; **a Lei Geral de Proteção de Dados pessoais, a contribuição da sociedade civil foi fundamental, não só das entidades de defesa do consumidor, de direito da comunicação, mas por exemplo, nas discussões da LGPD, no final de tudo, houve uma articulação grande das entidades desses setores com entidades empresariais pra se chegar a um consenso**. Eu acho que a gente amadureceu nesse debate, tudo isso com **contribuição do Comitê Gestor da Internet no Brasil** (grifo nosso).

Em complemento, Antonio Cavalcanti, do setor governamental, também comenta sobre tais marcos regulatórios, ao pontuar que:

A participação do terceiro setor e da academia são fundamentais. Basta olharmos a contribuição dada por eles ao longo da história do Comitê Gestor da Internet no Brasil - CGI.br, na construção

35 “A Coalizão Direitos na Rede é uma rede de entidades que reúne mais de 50 organizações acadêmicas e da sociedade civil em defesa dos direitos digitais, tendo como temas principais de atuação: acesso, liberdade de expressão, proteção de dados pessoais e privacidade na Internet” (informação coletada do website <https://direitosnarede.org.br/quem-somos/>)

do Marco Civil da Internet, da Lei Geral de Proteção de Dados – LGPD e, mais recentemente, no processo legislativo do PL das Fake News (PL 2630/2020).

Guilherme Araujo, do setor privado, por sua vez, comenta em acordo, ao comentar que o PL 2630/20 foi uma mobilização que nasceu da sociedade civil com foco principal em moderação de conteúdo, apesar de ter aglutinado outros temas ao longo do tempo.

No que diz respeito à postura do terceiro setor e academia enquanto atores que exercem pressão diante da atuação dos setores privado e governamental, Sthephany das Neves, membra da academia, afirma que tais se dedicam a acompanhar o que é feito no campo da moderação de conteúdo pelas plataformas, Assim, conseguem exercer um papel de controle do interesse público, seja por meio da demanda por maior transparência e/ou verificação dos procedimentos conforme parâmetros legais.

Nesse sentido, Theo Freitas, membro do terceiro setor, realiza ponderações sobre o papel de pressão exercido por tais setores e também de auxílio para com o setor privado:

Na academia e sociedade civil, qual é o papel, é... **eu acho que é duplo. Há um papel de avaliar e pressionar as empresas, para que elas fiquem melhores nessa atividade, para que elas sejam mais transparente em relação a essa atividade**, para que direitos não sejam, né... a academia, e a sociedade civil. tem... que ser o grilo falante, vamos dizer assim, consciência, que vai apontar as contradições entre os modelos de negócios e as soluções implementadas, entre os direitos em jogo e os modelos de negócio, fazendo essa **triangulação**. Então tem esse lugar de grilo falante, mas eu também acho que tem um lugar de **ajudar o setor privado a informar a sociedade sobre essa atividade**, porque a sociedade é muito mal informada sobre essa atividade, né. É... as pessoas analisam a moderação de conteúdo com base em casos individuais de remoção... (grifo nosso)

Apesar da quase totalidade dos entrevistados terem opinado em prol da atuação positiva do terceiro setor e academia, diversas críticas foram apresentadas às atividades de ambos, com ênfase no terceiro setor. Nesse sentido, as principais endereçadas a tal agente foram: atuação limitada em termos de escassez de recursos, difícil acesso à informações sigilosas das plataformas, baixa qualificação em relação a temas específicos e apresentação de posições enviesadas - visto que muitas organizações do terceiro setor são financiadas por grandes plataformas. Em relação

à academia, foi apresentada uma crítica pontual em face à sua postura isolada e pouco proativa.

Uma crítica também apresentada por alguns entrevistados foi quanto às organizações ou pessoas que, em decorrência do fluxo de informações e estrutura da internet atual, se consideram especialistas momentâneos de temas em alta, quando, na verdade, **não possuem a devida expertise para falar sobre tais conteúdos (4x)**. Nesse sentido, Theo Freitas pontua:

O problema é quando o terceiro setor participa sem, sem... estudar, vamos dizer assim. Mas eu não acho que seja a regra, que aí não dá nem pra falar que é terceiro setor direito, aí é, sei lá, um achismo de alguém. Eu não acho que existam grupos organizados que falam sobre moderação de conteúdo hoje que não saibam absolutamente o que eles estão falando como regra, isso é exceção.

Caio Fernandes, membro do setor governamental, se posiciona em acordo a esse pensamento. Assim, ao comentar sobre o tópico, ressalta que a possibilidade de se expressar foi uma conquista da Internet, a qual é uma iniciativa coletiva e colaborativa - elementos inerentes à estrutura multissetorial da Governança da Internet. No entanto, o que observa é que, muitas vezes, grupos de pessoas participam do debate no espaço virtual de maneira pouco produtiva, gerando ruídos e “se deixando deslumbrar pelo poder de voz”. Assim, geram prejuízos para o ambiente virtual e para a sociedade como um todo. Demarca, por sua vez, que nessa fala não se refere a organizações especializadas que se dedicam ao estudo de temáticas específicas, mas a um público geral observado em alguns espaços.

No que diz respeito às limitações enfrentadas pelo terceiro setor e academia, um dos posicionamentos apresentados foi o de Lucas Oliveira, membro do terceiro setor, que comenta sobre como o difícil acesso a algumas informações do setor privado podem comprometer a atuação do terceiro setor e academia, uma vez que se torna difícil “decifrar a caixa preta da moderação de conteúdo”. Logo, configura-se um cenário problemático em que não é possível propor mudanças estruturais, visto que os procedimentos adotados não são publicados. Um exemplo dado nesse sentido é quanto aos critérios utilizados para definir quando a moderação é feita de forma automatizada e quando é feita de forma manual.

Nesse sentido, Ana Carvalho, membra do terceiro setor, abrange uma série de limitações em sua fala, ao dizer que:

A comunidade técnico científica é **a que tem mais a oferecer nesses debates, mas é a menos ouvida**, junto com a sociedade

civil. São as **partes mais fracas, são as partes que têm menos influência, são as partes com menos dinheiro**, para conseqüentemente ter mais poder de influência... e aí... ao mesmo tempo, a sociedade civil é a parte que vai ser mais afetada por esse processo de regulação (grifo nosso).

Uma limitação apresentada de maneira pontual por um entrevistado foi a respeito do caráter enviesado da atuação do terceiro setor. Nesse sentido, Vicente Moura, membro do setor governamental, ressalta que a grande maioria das organizações que compõem a sociedade civil são financiadas pelas plataformas. Assim, em sua opinião, mesmo que as pesquisas não apresentem um direcionamento explícito, existe uma espécie de efeito inibitório (*chilling effect*) quanto às publicações e objetos de pesquisa. Esse movimento seria então um lobby indireto e “extremamente eficaz”, o qual também ocorreria na captação de membros da academia e sociedade civil organizada com financiamentos direto de pesquisa, impedindo assim um posicionamento mais duro diante dos processos legislativos e mesmo das práticas adotadas para moderação de conteúdo pouco transparentes. Diante disso, Vicente opina que a sociedade civil deveria buscar meios mais independentes para a realização das pesquisas, apesar de não ver isso ocorrendo em um futuro próximo.

Outro tópico contemplado pelos entrevistados foi quanto à forma como o terceiro setor e a comunidade técnico-científica se mobiliza diante das temáticas de interesse. Nesse sentido, alguns opinaram no sentido de que tais setores apresentam caráter reativo, se **mobilizando apenas diante de episódios focais (7x)** o que, por sua vez, comprometeria a qualidade de suas ações de maneira sistêmica (visto que alguns temas seriam super considerados e outros, desconsiderados). Nesse sentido, Augusto da Cruz, membro do setor governamental, atrela esse movimento ao funcionamento das redes sociais, as quais conferem maior visibilidade e engajamento às pautas em alta. Assim, diz que “o próprio Twitter vai dizer, em certos momentos essa coisa vai ser um trend topic e em algum momento todo mundo entende que alguma coisa deveria ter sido bloqueada ou então alguém entende que não.” Dessa maneira, por mais que as organizações adotem posturas mais vigilantes diante de outras temáticas, a população como um todo irá se interessar e dar visibilidade a poucas pautas.

Theo Freitas, por sua vez, apresenta posição semelhante, defendendo o seguinte:

as pautas ganham tração em apoiadores, em redes sociais... sabe, **a coisa consegue avançar quando há casos emblemáticos assim, o que é uma... tristeza, porque eu acho que as entidades da comunidade civil se esforçam para puxar esse assunto fora da lógica, vamos dizer assim, só de... discutir um caso individual.** Mas o ponto é que, elas ganham atenção e se mobilizam quando isso **vai pra pauta** (grifo nosso).

Samuel Cardoso, membro do terceiro setor, comenta sobre essa limitação do terceiro setor da seguinte maneira:

(...) do Terceiro Setor, a gente vê uma atuação positiva, mas que também acaba sendo limitada por conta da **profusão de temas**. Então a gente acaba tendo **muita coisa pra tomar conta, de maneira geral, e poucas organizações, no limite, né, em cada contexto nacional, a gente tem um número de questões que extrapola a capacidade de atuação dessas organizações do Terceiro Setor, que acabam priorizando aquilo que tá na agenda do momento**, os problemas que estão sendo criados pelas empresas de tecnologia ou pelos próprios tomadores de decisão e aí uma falha que existe na proposição de agendas positivas, né? Agendas que proativamente incidam sobre essas questões de moderação, né? Então, pra dizer uma falha, por exemplo, desse aspecto, **existem algumas regras no Marco Civil da Internet que poderiam já ter sido aprimoradas ou complexificadas, mas que de maneira geral nesses oito anos, quase dez anos, o Terceiro Setor não foi muito proativo em levantar essas questões e promover os debates e sugerir os aprimoramentos que seriam necessários**. Ele acaba, nesse ponto de vista, negativo, atuando muito **a reboque da situação e pouco com a definição de uma agenda prévia do que precisa ser feito** (grifo nosso).

Diante desse cenário, Vicente Moura, que caracteriza a sociedade civil como “reativa”, comenta que sente falta de uma incidência mais propositiva nos processos legislativos, “efetivamente propondo soluções, ainda que pragmáticas, de elaboração de dispositivos”. Complementa assim, que considera a pesquisa científica, relatórios e análises críticas sobre eventuais processos de regulação pouco práticos.

Por fim, Maria Teixeira elogia o terceiro setor em suas ações e comenta sobre a questão da mobilização sob um ponto de vista distinto. Para a entrevistada, o foco em grandes temas garante uma maior concentração de forças e, logo, uma mobilização mais coesa, reunindo mais grupos que fazem parte da sociedade civil organizada. Ademais, comenta positivamente sobre a capacidade do terceiro setor em se inserir e ter voz nos espaços institucionais.

4.3. Sobre a avaliação dos diferentes modelos regulatórios

Neste tópico, os entrevistados apresentaram a sua opinião acerca dos diferentes regimes de moderação, conforme questionado nas perguntas 8 a 12 e 14.

Sobre a autorregulação por parte das plataformas na moderação de conteúdo

No que diz respeito à autorregulação exercida pelas plataformas, foram consideradas as regras de moderação de conteúdo criadas por elas próprias. Assim, quando questionados sobre a avaliação de tais diretrizes em face à adequação aos Direitos Humanos, os entrevistados se dividiram nas seguintes opiniões: consideraram a aplicação **predominante positiva(7x), positiva e necessária(1x), predominantemente necessária (7x), negativa (2x)** e um entrevistado as classificou como apresentando elementos positivos e negativos. Dentre tais, 4 entrevistados apontaram dificuldades na aplicação dos parâmetros de Direitos Humanos, especificamente, e 5 entrevistados pontuaram dificuldades gerais.

Dessa forma, em face aos argumentos apresentados em prol da atuação positiva das regras aplicadas pelas próprias plataformas, tais foram: garantia de **liberdade e autonomia às plataformas(4x)**, possibilidade de adequação às eventuais **novas demandas(2x)**, manutenção de segurança ao ambiente online e garantia de comprometimento por parte das plataformas. Nesse sentido, Igor Peixoto, membro do setor empresarial, defende a existências de tais regras da seguinte forma:

eu acho que faz sentido as plataformas criarem suas próprias regras, pra começar, porque elas têm **naturezas diferentes, elas têm propostas diferentes**, as grandes de uma certa forma são globais, mas se a gente pensar em plataformas no geral podem ter **ambientes geográficos diferentes**, elas também têm **sistemas de enforcement diferentes**, então elas têm práticas comuns que **combina o uso de inteligência artificial com a atuação de moderadores de conteúdo**, mas esses sistemas, eles acabam no detalhe sendo diferentes, porque são empresas diferentes (grifo nosso)

Em complemento, Guilherme Araújo, membro do setor privado, pontua que é fundamental que as plataformas tenham suas próprias regras para a manutenção da segurança nos ambientes virtuais, frisando que devem sempre ser atualizadas. Ademais, para o entrevistado, a autorregulação é uma maneira de poder se adequar às novas demandas. Assim, como exemplo, cita que há alguns anos nenhuma

plataforma se preocupava em desenvolver políticas quanto à desinformação, o que poderia ser inclusive visto como “querer arbitrar o que é verdade ou não”. No entanto, hoje em dia, frente a questões recentes voltadas à saúde pública e a contextos eleitorais, as plataformas incorporaram essa preocupação, desenvolvendo diretrizes específicas para o assunto.

Em sentido semelhante, Sofia Pires, membra do terceiro setor, comenta sobre as respostas rápidas que as plataformas costumam oferecer em eventos de caráter urgente ou emergencial, citando como exemplo a pandemia do coronavírus e a Invasão do Capitólio, nos Estados Unidos.

Já no que diz respeito às dificuldades gerais apontadas pelo entrevistados quanto à autorregulação das plataformas, destacam-se: **opacidade (4x)**, **seletividade na aplicação (3x)**, dificuldades na **implementação e acompanhamento das regras pelas empresas (3x)**, adequação regulatória ou “**controle da legalidade**” (3x) e adequação ao interesse público.

Dessa forma, no que diz respeito à demanda por mais transparência, Samuel Cardoso, membro do terceiro setor, argumenta que a criação de regras pelas próprias plataformas é uma necessidade; no entanto, tais precisam ser mais transparentes e melhor divulgadas. Assim, complementa dizendo que “na nossa visão brasileira de direito do consumidor, deveriam ser **prestadas por meio de informações adequadas, evidentes e inteligíveis para a população em geral, o que não é o caso** (grifo nosso).” Em acordo, Ana Carvalho, do terceiro setor, também faz referência a um mecanismo legal para exemplificar como a autorregulação poderia ser mais transparente:

mas eu acho que elas devem ser **mais explícitas e mais transparentes**, em que sentido? que eu estava falando, é... devia ser como um Código Penal, digamos assim, tipo... **você fez isso... sua sanção vai ser essa. Se você reincidir nessa conduta, sua sanção vai ser essa...** Sabe? Porque eu acho que é isso que faz falta, assim, nas atividades de moderação de conteúdo que a gente vê hoje, é... isso não tá claro (grifo nosso).

No que diz respeito à questão da seletividade na aplicação das regras por parte das plataformas, Sofia Pires, membra do terceiro setor, defende que existiriam dois tipos de aplicação: “(...) um tipo de aplicação de regras pras **democracias ou espaços que importam**, e um tipo de aplicação de regras para **democracias e espaços que importam menos** (grifo nosso).”

Nesse sentido, comenta que as revelações da ativista Frances Haugen, ex-funcionária do Facebook, no ano de 2021, foram importantes para ilustrar como as plataformas

“prestam atenção em níveis diferentes para lugares diferentes do mundo”. Assim, isso impactaria na maneira como a moderação de conteúdo acontece em espaços do Sul Global *versus* Norte Global, a depender do nível de prioridade dado por cada plataforma.

Maria Teixeira, membra da comunidade técnico-científica, apresenta posição diferente de Sofia ao classificar a autorregulação pelas plataformas como negativa. No entanto, as duas entrevistadas concordam no que diz respeito à seletividade na aplicação das regras e consequente ausência de parâmetros definidos. Nesse sentido, diz:

(...) E eu acho que elas **não são efetivas**. Para essa criação de regras de moderação. É isso, **eu não acredito nessa autorregulação das plataformas porque pra mim tudo parece dois pesos e duas medidas**. Não é que ele vai tratar um caso de racismo igual ao outro, ele vai **particularizar**. Se for um agente político que expressar racismo nas redes sociais, será que ele vai ser punido, será que ele vai ser moderado, será que ele vai ser banido imediatamente, aquele conteúdo? Já vi várias vezes que não (grifo nosso).

Em complemento à fala da entrevistada, Emanuella Nogueira, membra do terceiro setor, defende a aplicação de tais regras a partir de padrões internacionais de Direitos Humanos, o que seria uma possível forma de criar um maior controle diante do mal estabelecimento de mecanismos de regulação observado pelas plataformas.

Quanto às dificuldades apontadas pelos entrevistados no que diz respeito à identificação de parâmetros de Direitos Humanos, aqueles que se manifestaram nesse sentido defendem que, por questões culturais e sócio-políticas, definir o que seria de fato uma regra que age em consonância a tais princípios é um desafio. Visto isto, as plataformas precisam de se atentar a essa diversidade, considerando sua atuação por vezes global. Nesse sentido, Vicente Moura, membro do setor governamental, comenta:

Então, é, se você pergunta especificamente sobre a **efetividade**, eu acho que, por mais que exista uma literatura, um consenso internacional sobre alguns direitos humanos, existe uma **necessidade de particularização de cada contexto cultural, nacional, eventualmente até mesmo regional, em relação a essas formas de moderação de conteúdo, em torno de direitos fundamentais específicos**, né, e as plataformas têm capilaridade, interesse tem lucrado muito com esses usuários regionalizados, territorializados, mas na hora de fazer a particularização do seu

modelo de negócio pras necessidades dessas pessoas, elas se recusam, falam que é muito custosa ou trabalhosa, então eu acho que tá na hora de essa desculpa não ser mais aceita (grifo nosso).

Em complemento, Igor Peixoto, membro do setor privado, apresenta um posicionamento em acordo, ao dizer que, ao traçar as regras próprias de moderação de conteúdo, as plataformas pensam nas “**necessidades que elas têm de tornar o espaço minimamente saudável**, aquela coisa da moderação enquanto produto, mas essa compatibilização com o direito internacional dos direitos humanos, ela **nunca foi uma prioridade das plataformas** (grifo nosso).”

Sobre a regulação estatal na moderação de conteúdo

Como regulação estatal, compreende-se aqui a definição e a aplicação de regras por autoridades estatais. Assim, quando questionados sobre suas posições a respeito das normas emitidas por autoridades estatais para a adequação da moderação de conteúdo aos direitos humanos, a maioria dos entrevistados se posicionou predominantemente **a favor desse tipo de regulação(11x)**, parte se apresentou **predominantemente contra(4x)**, dois entrevistados não se posicionaram e um apresentou argumentos tanto a favor como contra.

Dessa maneira, tem-se que o ponto mais recorrente utilizado para defender a atuação positiva da regulação estatal foi o **estabelecimento de parâmetros mínimos(5x)** para a definição das regras de moderação de conteúdo, traçando assim “mínimos e máximos” não existentes no cenário brasileiro atualmente. Em conjunto, foram também citados: o estabelecimento de parâmetros em acordo aos Direitos Humanos, a adequação/pressão à atividade das plataformas e a importância no estabelecimento de diretrizes específicas. Ademais, todos os entrevistados que se posicionaram a favor da regulação estatal o fizeram defendendo seu caráter positivo a partir da necessidade. Nesse sentido, Samuel Cardoso, membro do terceiro setor, comenta

É, eu acho que a legislação, ela é necessária para servir como parâmetro para essas várias normas internas das plataformas, estabelecer o mínimo **pra que justamente essa defesa dos direitos não seja estabelecida como uma vantagem concorrencial ou como um recurso adicional para quem paga pelo serviço ou para quem contrata um serviço especificamente**, né? Então, claro que a gente tem, acho que a legislação tem que estabelecer esses **patamares mínimos a partir dos quais as empresas possam promover os seus serviços**, né? (...)

Em complemento, Ana Carvalho, membra do terceiro setor, considera ser preciso haver uma regulação estatal que “diga o mínimo, é... que diga exatamente isso, que

os dispositivos, as regras das empresas, tem que ser... claras, tem que ser explícitas, tem que ser transparentes, tem que tá lá escrito dizendo o que que é e o que não é.” Essa fala encontra acordo na proferida por Maria Teixeira, membra da comunidade técnico-científica, a qual afirma que tais normas são fundamentais e devem estar “escritas, tem que estar no papel, tem que estar aprovado no Congresso Nacional. O Judiciário podendo julgar as coisas de acordo com essa legislação.”

Luiz Porto, por sua vez, comenta sobre a necessidade da atuação estatal por meio da regulação ao dizer:

o que tá faltando é isso, nas políticas de plataforma, uma **harmonização por força de lei**, já que as plataformas por si só, apesar dos princípios importantes, os importantes princípios de Santa Clara nesse campo, da perspectiva da autorregulação, da perspectiva da indústria, **não serem suficientes para criar dentro do ambiente regulatório, e aí sim, privado, corporativo, da regulação privada transnacional, diretrizes e regras uniformes a respeito da moderação de conteúdo** (grifo nosso).

Observa-se assim que as manifestações em vista ao estabelecimento de parâmetros mínimos visam justamente aos campos em que há um “vácuo regulatório”. Nesse sentido, os entrevistados que defendem esse ponto de vista enxergam nos órgãos estatais a competência e legitimidade necessária para a definição desses padrões que hoje não existem no cenário da moderação de conteúdo. No entanto, fato é que tal legitimidade não foi unânime entre os entrevistados, visto que aqueles que se posicionaram contra a regulação estatal apresentaram um número maior de críticas do que os que se posicionaram a favor apresentaram como pontos positivos.

Nesse contexto, os argumentos apontados em prol do caráter negativo das regras criadas pelos atores estatais foram: falta de conhecimento técnico por parte dos agentes estatais, falta de diálogo com os demais setores no estabelecimento de regras, não aplicação de parâmetros internacionais (mas apenas internos), aplicação de parâmetros internacionais com foco na remoção como única solução, possibilidades de censuras, possibilidade de abusos na aplicação de regras, possibilidade de afetar a concorrência, flexibilidade e autonomia das empresas e a possibilidade de imposição de mecanismos legais de caráter impositivo.

Visto isso, Bárbara Silveira, membra do setor empresarial, defende sua posição:

Apoiamos uma abordagem prospectiva da regulamentação, considerando o impacto a longo prazo no ecossistema digital mais amplo, que protege a Internet Aberta e o acesso universal. Por isso, **acreditamos que qualquer tentativa de legislação que**

apenas reforce a posição dominante de empresas maiores prejudicará a Internet Aberta, a inovação e a escolha do consumidor de forma irreparável. Em suma, consideramos essenciais a **capacidade de concorrência robusta e a garantia de um fair playing feld entre os atores** (grifo nosso).

Nesse sentido, Guilherme Araújo, membro do setor privado, também apresenta receios diante de tal regulação estatal, ao dizer que “é preciso tomar cuidado” para que as normas em questão não resultem em uma forma de censura estatal ou pressão estatal por censura privada.

No que tange aos desafios da regulação por parte dos atores estatais, um ponto de vista foi expresso por três entrevistados de maneira semelhante. Trata-se da dificuldade de se regular a área da tecnologia, visto a “velocidade das inovações tecnológicas” e o “risco de rápida obsolescência” às legislações, quando muito detalhadas, de acordo com Antonio Cavalcanti, membro do setor governamental. Em acordo, Vicente Moura, também membro do setor governamental, diz:

Existe uma dificuldade muito grande em regular tecnologia, assim, então o trabalho legístico em torno dessa atividade, **ele deve ser flexível o suficiente pra perdura ao longo do tempo**, ou seja, pensar numa certa **vitalidade desse estatuto**, dessa lei ou desses regulamentos ao longo do tempo, **para se adaptar às novas tecnologias ou eventualmente ser alterado na medida em que essas tecnologias também apresentam novos desafios e isso é muito difícil.** O processo legislativo é difícil, ele é demorado; o processo regulatório, pra ser um processo regulatório bom, ele envolve **relatórios de impacto, ele envolve tomada de subsídios da sociedade, consulta pública, isso tudo são processos demorados**, assim, então eu acho que é um desafio insuperável que as formas de regulação estatal sempre vão estar atrás dos desafios propostos por essas tecnologias (grifo nosso).

Tendo em vista o comentário do entrevistado, é possível refletir sobre como as respostas e recursos disponíveis pelo Direito, muitas vezes são morosas, pouco efetivas e mesmo ineficientes. Em contraste à velocidade do fluxo de informações na rede e à quantidade de questões conflitantes que podem vir à tona, a possibilidade de regular precisa de passar por uma série de etapas, envolvendo jogos de interesse e desafios diversos. Diante disso, no cenário da Governança da Internet, a regulação de plataformas por órgãos estatais é um tema que encontra opiniões distintas.

Nesse cenário, Lucas Oliveira, do terceiro setor, comenta que uma regulação

estatal, quando de caráter impositivo, pode afetar o ecossistema das plataformas de maneira danosa no que diz respeito às regras sobre moderação de conteúdo, visto que as plataformas devem ter sua liberdade e autonomia para determinar como os conteúdos serão moderados. Nesse tema, afirma:

(...) ela [a internet] é plural e é bom que ela seja assim, então cada plataforma tem habilidade de atrair certos conteúdos e se desvincular de outros conteúdos para criar uma determinada comunidade focada em um determinado assunto e tema.

Em complemento, o entrevistado criticou a regulação estatal por meio da criação de mecanismos com listas exaustivas do que poderia ou não ser moderado, citando a Medida Provisória 1.068 de 2021³⁶ como próxima dessa lógica. No caso, a referida norma, dentre outras previsões, trazia um rol de conteúdos que autorizariam sua exclusão, suspensão ou bloqueio por parte da plataforma.³⁷ Apesar disso, Lucas defende a aplicação de regras estatais que implementem maior transparência às práticas de moderação de conteúdo, proteção do interesse público na esfera digital e exigência do respeito ao devido processo.

Sobre as recomendações sem efeito legal (Soft Law)

No presente tópico são abordadas as opiniões dos entrevistados quanto ao impacto das recomendações sem efeito legal (as chamadas Soft Law) diante da adequação das práticas de moderação de conteúdo aos direitos humanos. Assim, como exemplo ilustrativo, foi mencionado na pergunta os Princípios de Santa Clara, os quais estabelecem padrões de transparência e *accountability* para práticas de moderação de conteúdo.

Ao se posicionarem, um pouco mais da metade dos entrevistados classificou tais mecanismos como **positivos e eficientes (9x)**, sendo que 3 desses entrevistados apresentaram críticas - as quais não colocaram o caráter da eficiência em questão. Em seguida, as classificações foram: **positivos, mas ineficientes (4x)**, **apenas ineficientes (2x)**, **negativos (1x)** e **sem posição (2x)** - um dos entrevistados preferiu se omitir, visto que não se considerou na condição de responder a pergunta por saber pouco sobre o tema.

36 BRASIL. Medida Provisória nº 1.068, de 6 de setembro de 2021. **Altera a Lei nº 12.965, de 23 de abril de 2014, e a Lei nº 9.610, de 19 de fevereiro de 1998, para dispor sobre o uso de redes sociais.** DF: Presidência da República, [2021]. Disponível em: <https://www.in.gov.br/en/web/dou/-/medida-provisoria-n-1.068-de-6-de-setembro-de-2021-343277275>. Acesso em: 02 dez. 2022.

37 Alguns dias após a publicação da norma, a ministra-relatora Rosa Weber, do STF, deferiu pedidos de tutela cautelar em 7 ações diretas de inconstitucionalidade contra a MP 1.068/21, para suspender os seus efeitos. Ver mais em: CARNEIRO, Luiz Orlando. Os argumentos de Rosa Weber para suspender a MP que dificultava remoção de conteúdo. **JOTA, 15 de setembro de 2021.** Disponível em: <https://www.jota.info/stf/do-supremo/os-argumentos-de-rosa-weber-para-suspender-a-mp-que-dificultava-remocao-de-conteudo-15092021>. Acesso em: 04 jan. 2022.

Dessa forma, os principais argumentos apresentados em prol do caráter positivo e/ou eficiente de tais mecanismos, foram: auxílio na construção de **parâmetros para a criação de regras de moderação de conteúdo (5x)**, colaboração para a **criação de diretrizes regulatórias (3x)**, execução de **pressão para com as plataformas (4x)**, criação de espaços de debate entre vários setores e construção de fontes de informação relevante.

Dessa forma, no que diz respeito à criação de parâmetros, os entrevistados utilizaram diferentes expressões de sentido semelhante, a exemplo de: criação de molduras, indicação de caminhos e ajuda em nortear os trabalhos. Nesse sentido, Guilherme Araújo, membro do setor empresarial, afirma que “(...) toda boa construção parte do estabelecimento de princípios que às vezes vão vir da academia, às vezes da sociedade civil, às vezes das empresas.... mas numa **gama de princípios compartilhados** (grifo nosso)”. Tal opinião encontra acordo na fala de Antonio Cavalcanti, membro do setor governamental, ao afirmar que “instrumentos desse tipo ajudam a moldar mentes e corações e **preparam o terreno político para medidas com fundamento legal e protetivas dos direitos humanos mais à frente** (grifo nosso)”. Já para Igor Peixoto, membro do setor privado, tais recomendações funcionam como “sinalizadores de expectativas de setores impactados pelas plataformas”, permitindo assim ao setor privado uma melhor compreensão por parte dos demais agentes.

Assim, por mais que os mecanismos de Soft Law não apresentem força normativa, são vistos por alguns entrevistados como instrumentos importantes na criação de marcos regulatórios, ao passo que se tornam inspirações nos processos legislativos. Dessa maneira, ao determinarem princípios básicos e cenários ideais, fundamentam a base normativa pretendida em acordo aos direitos humanos e éticos. Nesse sentido, Luiz Porto, membro da comunidade técnico-científica, os considera como “mais abertos às transformações”, afirmando também que “é mais fácil atualizar e modernizar princípios de recomendações e diretrizes hoje que capturem muito mais a essência das transformações.” Dessa forma, o entrevistado considera os mecanismos de Soft Law como positivos e eficientes, vendo-as como motivadoras de decisões judiciais e empresariais. Como ilustração, comenta sobre o impacto que as diretrizes e recomendações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) têm alcançado no cenário brasileiro.

No que tange à pressão que as recomendações sem efeito legal proporcionam nas empresas, Theo Freitas, membro do terceiro setor, comenta:

As empresas estão se conformando nesse sentido, vamos pensar que o Board é um exemplo. Isso quer dizer que as empresas vão implementar exatamente o que essas regulações querem? Não. Vamos pegar, por exemplo, o último relatório de transparência do Facebook. É... comentários que eles fizeram sobre erros, as

demonstrações sobre a mudança nos números de remoção por categoria, e tal. **Ele cria esses esforços nesse sentido. Isso não é algo proporcionado por regulação, e também não é um negócio que as plataformas tão simplesmente fazendo porque elas acham legal, elas fazem por pressão. E a pressão se dá a partir desses instrumentos**, então... eu acho que, em alguma medida, surte efeito sim (grifo nosso).

Em acordo, Antonio Cavalcanti pontua que, embora tais instrumentos não possuam poder coercitivo, desempenham papel relevante ao provocar efeitos diretos e indiretos nos mecanismos de autorregulação das plataformas, visto que “ajudam a criar um ambiente que induz e pressiona por práticas mais humanas no uso das mais variadas tecnologias.”

Já quanto aos argumentos apresentados em justificativa ao caráter negativo e/ou ineficiente dos instrumentos em discussão, todos, direta ou indiretamente, disseram respeito ao caráter limitado que os instrumentos possuem em face aos interesses comerciais das plataformas, tendo em vista o modelo de negócios vigente. Assim, Vicente Moura, membro do setor governamental, comenta nesse sentido:

Eu acho que elas são positivas, mas eu sou meio cínico, assim, eu acho que o interesse **econômico superveniente a qualquer tipo de acordo, declaração de intenções ou até mesmo práticas, códigos de conduta setoriais que essas empresas venham a desenvolver**, elas não vão fazer algo que vai contra, por exemplo, os interesses econômicos dela. E moderação de conteúdo é problemática justamente porque **um dos principais interesses econômicos dessas empresas é ganhar engajamento de usuários e engajamento normalmente ocorre por meio de conteúdos polêmicos, de notícias falsas, de conteúdos não moderados**, assim, então eu não vejo como **princípios não vinculantes ou normas voltadas para ética de uma empresa** seriam capazes de propor medidas efetivamente capazes de regular a moderação de conteúdo, **então eu acho que elas não são suficientes** (grifo nosso).

Em contexto semelhantes, Lucas Oliveira, membro do terceiro setor, considera as recomendações sem efeito legal positivas, mas “com efeito moderado, para não dizer nulo na moderação de conteúdo das plataformas digitais”. O que leva a isso, em sua opinião, é a falta de interesse apresentado pelo setor privado em realmente considerar tais instrumentos em suas ações. Como exemplo, cita que a ausência de referência às grandes cartas de princípios pelo Oversight Board em suas decisões.

Diante das limitações colocadas pelos entrevistados que consideram a ineficiência dos mecanismos de Soft Law, uma questão colocada foi quanto à necessidade de força legal para que tais recomendação de fato sejam seguidas. Esse posicionamento é ilustrado por Fernanda Rezende, membra do terceiro setor, em seu comentário:

(...) se a empresa está assistindo a veiculação de conteúdos inequivocamente ilegais na sua rede, de práticas que põem - da utilização abusiva e discriminatória de dados pessoais - e a empresa não faz nada, a gente precisa ter mecanismos de responsabilidade. Porque sem isso elas não vão mudar a conduta delas. **A gente precisa de todo um sistema que integre essas soft laws até e as regras das empresas e a atuação estatal. Construída com base no mecanismo multissetorial, como está proposto no PL 2630** (grifo nosso).

No que diz respeito ao entrevistado que considera as regulações sem efeito legal como negativas, seu argumento principal foi o de que tais instrumentos estariam mais direcionadas a promover um modelo de negócios do que para promover direitos humanos e garantias fundamentais. Nesse sentido, Samuel Cardoso, membro do terceiro setor, afirma:

Para a promoção dos direitos humanos em termos de eficiência, eu acho que elas ficam aquém, né, elas poderiam ser um instrumento de promoção de direitos humanos, mas elas acabam ficando abaixo do que poderiam e talvez eu pudesse avaliar que de maneira geral também **elas estão mais direcionadas a promover um modelo de negócio do que exatamente promover direitos humanos e garantias fundamentais**. Então, a gente tem diversos exemplos de situações em que, na medida em que a violação de direitos humanos ou a dúvida em relação à violação de direitos humanos promove atenção e essa atenção movimenta a economia da atenção, as plataformas acabam sendo mais permissivas para esse tipo de violação do que propriamente coercitivas ou limitantes desse tipo de violação (grifo nosso).

Dessa forma, tem-se em vista que o modelo de negócios das plataformas, alimentado pela busca por engajamento e lucro, configura-se como fator determinante no momento da (não) aplicação dos princípios determinados por tais instrumentos, na perspectiva do entrevistado.

Sobre normas estatais com diretrizes procedimentais

Acerca da criação de normas estatais com diretrizes específicas para moderação de conteúdo, **houve ampla manifestação positiva (14x)**, destacando a necessidade

de **bases mínimas para um procedimento ideal (5x)**. Nesse sentido, defendeu-se a criação de um **devido processo para a moderação (3x)** e o estabelecimento de garantias processuais mínimas aos usuários, como, por exemplo: **quais as hipóteses e o que será feito nas atividades de moderação, motivos e regras utilizadas para uma decisão (3x), possibilidade de recurso (5x)**, prazos para revisão de determinados conteúdos e regras para notificação de usuários que tiveram conteúdo moderado. A menção a essas situações demonstra uma preocupação em tornar mais compreensível o procedimento de moderação de conteúdo às pessoas que utilizam as plataformas, pensando, ainda, em maneiras de assegurar a contestação de decisões.

Para Luiz Porto, membro do setor técnico-científico, é necessário

fazer com que essas decisões [de moderação] estejam submetidas aos **controles judiciais e garantias processuais de usuários**, algo que não fica muito claro na atual sistemática das decisões de remoção por parte de algumas empresas e algumas plataformas, já que se nós observarmos nas políticas de comunidade, nas regras de comunidade ou nas políticas de plataforma não há previsão de qual procedimento o usuário tem que adotar, por exemplo, se precisar, se quiser contestar uma decisão de remoção.

[...] E aí sim, estabelecer, por exemplo, que na tomada de decisões de moderação de acordo com as políticas das plataformas adotadas **sejam empregadas as melhores garantias ou a garantia de proteção dos direitos e liberdades individuais dos usuários de internet, bem como observância das garantias processuais que lhes são asseguradas por força da Constituição, do Código de Processo e também dos tratados e convenções de que o Brasil é parte [...]** (grifo nosso).

Além disso, também foram elencados outros exemplos de possíveis diretrizes, como a determinação de: mecanismos a serem observados na remoção de conteúdo; criação de canais específicos para oferecimento de explicações sobre remoções; disponibilização das políticas de governança de conteúdo em linguagem acessível; exigência de **transparência em regras de moderação (3x)**, como por meio da **elaboração de relatórios de transparência (2x)** e o tipo de informação que devem conter; **divulgação dos parâmetros de tomada de decisão e critérios para moderação (2x)**, inclusive para equipes de pesquisa de universidade e institutos de pesquisa credenciados e certificados; incentivos a ações alternativas à remoção, como sinalização ou diminuição do alcance de conteúdo e checagem de fatos com rede de especialistas contratados; informações acerca do uso de inteligência artificial; avaliar o alinhamento do procedimento a parâmetros internacionais de direitos humanos;

respeito a direitos fundamentais, como a liberdade de expressão; e avaliação de desempenho das plataformas na aplicação dos seus termos de serviço, “de modo a proporcionar flexibilidade e reduzir os incentivos ao conteúdo excessivamente moderado” junto ao incentivo de investimento em soluções tecnológicas.

Entretanto, dentre os posicionamentos favoráveis, foram feitas algumas ressalvas acerca da criação desse tipo de lei. Pontuou-se a necessidade de se evitar o engessamento do procedimento de moderação, a fim de permitir sua evolução, assim como a preservação de certo grau de autonomia às plataformas para elaboração de seus termos de uso, em virtude de suas particularidades. Outro apontamento foi de que a criação de uma diretriz específica sobre o que e como moderar poderia criar apenas uma forma de fazer as coisas e que mudaria rapidamente.

Por outro lado, o reconhecimento de particularidades no funcionamento de cada plataforma serviu de argumento para um dos posicionamentos contrários a esse tipo de norma. Nesse caso, entendeu-se que caberia à plataforma estabelecer suas regras e, ao Estado, a exigência de transparência acerca delas. Outro posicionamento desfavorável reconheceu a possibilidade de diretrizes para determinar casos em que fosse possível o direito de resposta ou contestação, mas alegou que não caberia à lei dar prazo para a remoção de conteúdo.

Sobre normas estatais prevendo a remoção de conteúdos específicos

Esse debate esteve intimamente ligado com as considerações feitas pelos entrevistados sobre a viabilidade de a lei estabelecer casos específicos de conteúdos que deveriam ser removidos pela plataforma, à vista do que ocorreu com a edição da Medida Provisória 1.068/2021,³⁸ que trouxe um rol de hipóteses em que estaria autorizada a remoção. A pergunta feita demandava a atribuição de uma nota, de 0 a 10, acerca do quanto o participante se considerava favorável à criação desse tipo de norma, **sendo que a média obtida a partir de 14 pessoas que decidiram atribuir um valor foi de 4,8**. Essa nota, próxima de 5, esteve bem representada nos argumentos dos entrevistados, onde a maioria apresentou tanto considerações favoráveis como desfavoráveis.

Dentre os argumentos desfavoráveis, mencionou-se que estabelecer um rol específico de casos para remoção de conteúdo poderia gerar **incentivo a uma censura (2x)** e **remoção excessiva pelas plataformas (3x)**, assim como no caso de aplicação de penalidades administrativas significativas e estipulação de prazos curtos para

38 BRASIL. Medida Provisória nº 1.068, de 6 de setembro de 2021. **Altera a Lei nº 12.965, de 23 de abril de 2014, e a Lei nº 9.610, de 19 de fevereiro de 1998, para dispor sobre o uso de redes sociais**. DF: Presidência da República, [2021]. Disponível em: <https://www.in.gov.br/en/web/dou/-/medida-provisoria-n-1.068-de-6-de-setembro-de-2021-343277275>. Acesso em: 02 dez. 2022.

remoção. No ponto, pequenas empresas e novos serviços poderiam ser mais afetados por dispor de menos recursos para litigar ou pagar eventuais multas. Ainda, a criação desse tipo de rol poderia resultar em uma **censura por parte do próprio Estado (2x)**, como na hipótese de estados autoritários.

Aliás, a desconfiança em relação à forma com que o governo editaria uma norma nesse sentido apareceu mais de uma vez. Além da **falta de conhecimento técnico dos legisladores (2x)** e de uma **maior discussão no Brasil sobre os tipos de conteúdo que deveriam ser objeto de moderação (2x)**, o debate careceria de **definições mais específicas em torno de conceitos considerados “zonas cinzas” (2x)**, como no caso de liberdade de expressão, e alguns conteúdos, a depender do contexto em que publicados, poderiam ter significados diferentes, como é o exemplo de conteúdo de nudez. A falta de certeza de uma compreensão uniforme do Poder Judiciário, possível responsável pela supervisão dessas regras, acerca de temas como discurso de ódio, racismo, homofobia e outros também se mostrou um ponto de dificuldade.

Nessa direção, houve uma preocupação com a criação de categorias de conteúdos a serem removidos a partir de conceitos legais excessivamente abertos, como no caso desinformação e discurso de ódio, o que poderia gerar insegurança jurídica e impactos à liberdade de expressão. Assim, seria mais interessante focar na definição de conteúdos que já são considerados ilegais, **como exceções previstas em lei ou condutas já consideradas criminosas (9x)**. Alguns exemplos citados: pornografia infantil, pedofilia, *revenge porn*, direito autoral, crimes contra a vida e integridade física, apologia ao nazismo, racismo, dentre outros. Cabe destacar, porém, que um dos entrevistados sinalizou que crimes contra a honra não entrariam no rol de conteúdos a serem removidos, em razão da dificuldade que seria atribuída à plataforma em conseguir identificá-los.

Por outro lado, desinformação, assim como novamente racismo, homofobia, transfobia e violência contra a mulher foram observados por algumas das pessoas que defenderam a **aplicação desse tipo de norma para situações específicas ou pontuais (6x)**, como através de determinação judicial. Entretanto, também houve posicionamento a favor de um regramento estatal apenas em linhas gerais, isto é, uma parametrização sobre pontos inaceitáveis do ponto de vista social.

A criação de listas exaustivas de conteúdo pelo Estado poderia resultar no engessamento do processo de moderação e em impactos à pluralidade das plataformas, com a aplicação de uma única regra a todas elas. Dessa forma, a legislação não deveria apresentar um rol fechado de hipóteses. Além disso, ao focar no conteúdo, uma lei como essa também estaria **responsabilizando a plataforma pela ação de terceiros (2x)**, ameaçando, outra vez, direitos como a liberdade de expressão.

Para Bárbara Silveira, membro do setor empresarial, uma regulação deve

estabelecer padrões claros para os tipos de conteúdo que procuram abordar, com **definições e limites substantivos e consistentes com os padrões de direitos humanos**. Quando o conteúdo em questão é legal, mas um governo acredita que é necessário intervir, o marco regulatório deve fazer uma distinção clara entre esses tipos de conteúdo. **As solicitações do governo para a remoção de peças específicas de conteúdo com base na ilegalidade devem ser baseadas em um processo legal e fornecer transparência sobre como esses poderes são usados**. É uma questão fundamental de devido processo legal que um **órgão governamental competente**, e não um ator privado, seja responsável por determinar a ilicitude de um conteúdo. **As empresas devem ter liberdade para avisar as pessoas de que essa foi a base para a ação a ser tomada**. (grifo nosso).

A menção a um órgão governamental foi feita também por outro entrevistado, que referiu ser necessária uma maior fiscalização acerca da aplicação desse tipo de norma pelas plataformas, algo que poderia ser feito por uma agência similar àquela criada para fins de proteção de dados (ANPD), mas voltada a casos de moderação de conteúdo.

Nessa senda, outro entrevistado destacou a necessidade de se evitar a delegação de atividades estatais às plataformas, como a aplicação e fiscalização do cumprimento de lei, uma vez que configuraria uma espécie de privatização da jurisdição. Seria possível, nesse caso, realizar exercícios de impacto regulatório, a fim de “enfrentar coerentemente as formas e procedimentos de restrições desses direitos, mas também, ao mesmo tempo, demonstrar quais são os objetivos e o bem-estar social alcançado pela escolha regulatória de intervir no tema de moderação de conteúdo”.

Para além de obrigações, porém, dois entrevistados defenderam a criação de estímulos para que as plataformas realizem a moderação de conteúdo, como por meio da premiação daquelas que tenham bons trabalhos nessa área. Ainda, outro argumento destacou a necessidade de se pensar em uma base em torno de regras de transparência e devido processo antes de uma lei que determine a remoção de conteúdos específicos.

Por fim, um dos entrevistados pontuou que não caberia a remoção de conteúdo no caso de conversas privadas, hipótese em que o ajuizamento de processo contra o particular que proferiu ofensas seria o caminho mais adequado.

Sobre o regime de autorregulação regulada

Conforme já mencionado, o roteiro de perguntas foi elaborado com base em pesquisa prévia realizada a partir de revisão sistemática de literatura estrangeira e nacional acerca dos diferentes regimes de moderação de conteúdo. Nesse sentido, a pergunta sobre autorregulação regulada foi feita após a identificação do conceito na amostra analisada, de tal modo que se buscou compreender melhor se os entrevistados já conheciam o termo, o que entendiam como autorregulação regulada e como avaliavam esse modelo de regime.

Para **dois entrevistados, a autorregulação regulada seria, na verdade, uma espécie de correção**. Lucas Oliveira, membro do Terceiro Setor, afirmou que “o que existe é correção, é preservar a esfera de regulação, de autorregulação das plataformas digitais, e inserir o Estado no que lhe compete, por isso que correção e não autorregulação regulada”.

Por outro lado, para Theo Freitas, também membro do Terceiro Setor, a autorregulação regulada seria diferente da correção na medida em que, na primeira, o regulador apenas elaboraria diretrizes de atuação e realizaria a avaliação e validação do que será feito pelo regulado. A correção, de outra banda, exigiria a elaboração conjunta entre empresa e regulador do “caminho a ser percorrido” e dos detalhes, demandando um trabalho mais cooperativo.

A dificuldade em definir a autorregulação regulada como gênero ou espécie também foi encontrada na fala de Caio Fernandes, que afirmou ser difícil definir onde termina a autorregulação e onde começa a correção, uma vez que, no caso da autorregulação regulada, “você acena com a autorregulação ao mesmo tempo que você ameaça a correção”.

Apesar desses impasses, houve um certo consenso de que a autorregulação regulada seria a possibilidade de as próprias plataformas estabelecerem suas regras, mas **com base em diretrizes previstas pelo Estado (10x)**. No caso, o **regulador não estabeleceria o conteúdo dessa autorregulação (2x)**, mas haveria espaço para acompanhamento e métricas de avaliação da atuação das plataformas.

Para Theo Freitas, do terceiro setor, também caberia ao regulador validar a proposta de atuação das empresas, elaborada com base nas diretrizes estabelecidas, e, para Fernanda Rezende, do mesmo setor, a autorregulação regulada permitiria também a existência de uma instância de controle. Entretanto, para Stephany das Neves, membra do setor técnico-científico, esse modelo de regulação seria a possibilidade de potencialização dos mecanismos de autorregulação a partir de algum direcionamento legal ou *reinforcement* do Estado.

Com relação às percepções em torno da efetividade desse tipo de regime, alguns entrevistados se manifestaram no sentido de acreditar que a **autorregulação regulada poderia ser mais eficiente do que outros regimes (6x)**, havendo menção de que esta opção seria melhor do que uma regulação *top-down*, isto é, de cima para baixo, como é o caso de uma regulação puramente estatal. Entretanto, ressaltou-se que sua efetividade dependeria do próprio regulador, de uma boa aplicação e de que esse regime fosse tanto multissetorial como que tivesse uma perspectiva internacional integrada, a fim de não ser capturado ou burlado.

Outros participantes consideraram a autorregulação regulada um **caminho viável (2x)**, ressaltando-se novamente ser melhor do que uma regulação *top-down*, mas que deveria pautar por uma participação diversa, tanto no seu acompanhamento posterior como na formulação das diretrizes para as plataformas, envolvendo não apenas os setores governamental e empresarial. Para outros entrevistados, esse regime seria, na verdade, a **melhor solução que há no momento (3x)**, pois além de seguir uma tendência vista em outros países, de busca por uma correção, é uma forma de regulamentação que permite atuação estatal ao mesmo tempo em que assegura autonomia para as plataformas, sendo pública e privada.

Ainda, argumentou-se que o modelo de autorregulação regulada seria uma alternativa **melhor do que deixar apenas para as plataformas controlarem todos os seus processos (2x)**, incluindo a moderação, sendo necessárias outras instâncias para monitoramento da sua atuação. Para Antonio Cavalcanti, membro do setor governamental, esse é um regime que faz sentido diante do cenário atual do mercado de tecnologia:

(...) a mescla da autorregulação com uma regulação autocontida, feita pelo **Estado ou algum órgão regulador**, predominantemente principiológica e com perfil mais orientador e diretivo, **permite a manutenção do espaço de experimentação de negócios necessário para o desenvolvimento das plataformas.**

As características do mercado de tecnologia e o cenário de intenso e célere desenvolvimento de novos produtos e serviços pelas plataformas de internet **tornam a adoção dessa solução híbrida o cenário ideal.** (grifo nosso).

A menção a um órgão de controle também foi encontrada na fala de Fernanda Rezende, membra do Terceiro Setor, que mencionou novamente a importância de que esse espaço fosse multissetorial, a fim de ter participação dos diversos agentes da sociedade na construção da regulação. Outros dois argumentos destacaram, respectivamente, que a autorregulação regulada seria interessante por promover um local de diálogo entre o objeto de regulação e o regulador e que mecanismos de correção seriam o caminho adequado quando se trata de situações de risco a direitos fundamentais.

Entretanto, houve críticas à forma como esse regime foi traduzido no Projeto de Lei 2.630/2020, o chamado “PL das Fake News” que, apesar do nome e após uma série de modificações ao longo da tramitação no Legislativo, estava mais alinhado a uma regulação geral de plataformas. A proposta apresenta um modelo de autorregulação regulada que prevê a possibilidade de criação, por parte dos provedores, de uma “instituição de autorregulação voltada à transparência e à responsabilidade no uso da internet”, conforme se verifica abaixo:

Art. 35 Os provedores poderão criar instituição de autorregulação voltada à transparência e à responsabilidade no uso da internet, com as seguintes atribuições:

I – criar e administrar plataforma digital voltada ao recebimento de denúncias sobre conteúdos ou contas e tomada de decisão sobre medidas a serem implementadas por seus associados, bem como a revisão de decisões de conteúdos e contas, por meio de provocação por aqueles afetados diretamente pela decisão;

II – tomar decisões, em tempo útil e eficaz, sobre as denúncias e revisão de medidas abrangidos por esta lei; e

III – desenvolver, em articulação com as empresas de telefonia móvel, boas práticas para suspensão das contas de usuários cuja autenticidade for questionada ou cuja inautenticidade for estabelecida; (...)(grifo nosso) ³⁹

Além disso, o PL prevê atribuições ao Comitê Gestor da Internet, que teria o papel, por exemplo, de “realizar estudos, pareceres e recomendações sobre liberdade, responsabilidade e transparência na internet”⁴⁰. No ponto, a escolha por colocar o CGI **responsável por dialogar com o setor privado um parâmetro para os termos de conduta das plataformas (2x)** foi criticada pelos entrevistados, tendo sido apontado também que a proposta partiria do pressuposto de que toda moderação de conteúdo fere a liberdade de expressão.

Alegou-se que essa estratégia traria a interpretação de que a autorregulação regulada depende necessariamente de um órgão de revisão de decisões, quando, na verdade, bastariam normas, princípios e regras gerais de devido processo. Em razão disso, a previsão do PL seria pouco efetiva e frágil, com problemas de aplicação. No ponto, a desconfiança com relação a como o conceito de autorregulação regulada seria aplicada no Brasil também foi mencionada por outro entrevistado, que demonstrou preocupação com a distorção do termo.

39 BRASIL. Projeto de Lei 2.630 de 2020. **Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet**. DF: Senado Federal, [2020]. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2256735>. Acesso em: 19 dez. 2022.

40 BRASIL. Projeto de Lei 2.630 de 2020. **Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet**. DF: Senado Federal, [2020]. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2256735>. Acesso em: 19 dez. 2022.

Dúvidas foram igualmente levantadas acerca da efetividade de criação de políticas individuais por cada plataforma, sem que haja a harmonização das normas voltadas à moderação de conteúdo:

Agora, eu desconfio da efetividade de políticas regulatórias e individuais - desculpe, códigos individuais de autorregulação por parte das plataformas sem o potencial de harmonização e uniformização das soluções relacionadas à moderação de conteúdo, porque as plataformas teriam de fazer um esforço que elas hoje ainda timidamente fazem, que é poder, nos foros especializados, **sentarem com os especialistas e formuladores de política pra encontrar soluções harmonizadas e uniformizadas em tema de moderação**. Pergunta: isso acontece na atualidade? Não, se acontece é o próprio segmento da indústria, mas **o próprio segmento da indústria teria que então dar oportunidade pra outros atores multissetoriais entrarem e discutirem esse modelo, não é?** (grifo nosso).

Luiz Porto, membro do setor técnico-científico.

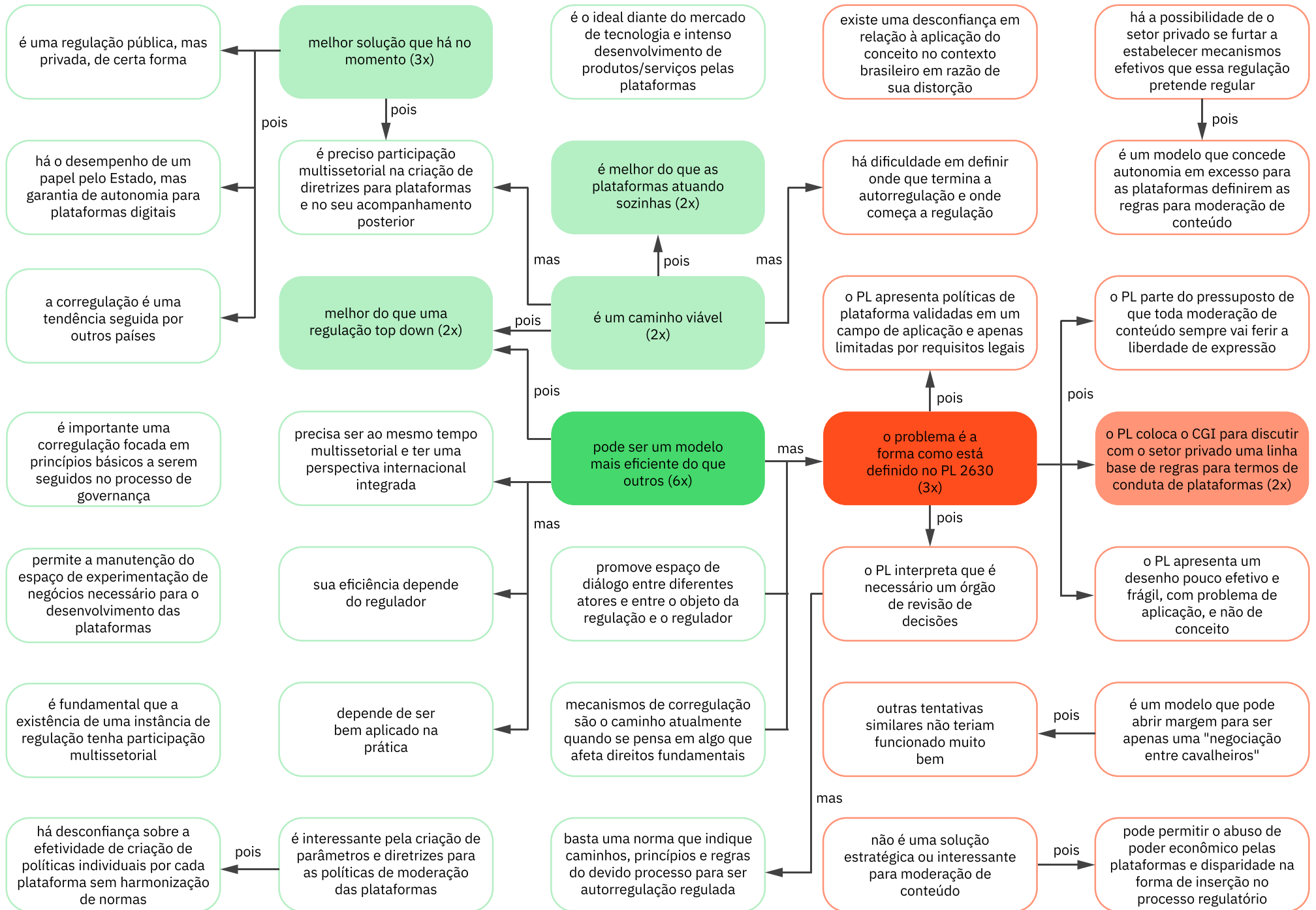
Nesse sentido, o entrevistado destaca que a autorregulação possui equívocos na forma como consta no referido projeto de lei, uma vez que aparentaria representar “um conjunto das políticas de plataforma validadas dentro de um campo ou um âmbito de aplicação”, isto é, como se pudessem ser limitadas somente por estritos requisitos legais.

Em relação às plataformas, ainda houve críticas quanto ao poder conferido às empresas por meio da adoção desse regime. Argumentou-se que ele não seria estratégico ou interessante para a moderação de conteúdo, diante da possibilidade de “abuso de poder econômico dessas plataformas, a disparidade da maneira como elas se inserem, inclusive, no processo de elaboração das leis atualmente, por meio de *lobby*”, assim como “de o setor privado se furtar a estabelecer, na prática, mecanismos efetivos que essa regulação pretende regular”.

Além disso, foi mencionado receio de que a autorregulação regulada, por meio da criação de um órgão regulador, pudesse não funcionar adequadamente, assim como no caso de outras agências, e que a sua atuação na regulação de plataformas pudesse ocorrer de forma pouco incisiva, abrindo margem para uma “negociação entre cavalheiros”.

Confira abaixo o mapa mental sintetizando os principais argumentos apresentados em relação ao modelo de autorregulação regulada:

Figura 3 - Principais argumentos positivos e negativos em relação ao modelo de autorregulação regulada



4.4. Sobre a regulação da moderação de conteúdo no Brasil

Este tópico aborda os argumentos apresentados pelos entrevistados acerca de sua avaliação sobre o modelo regulatório de moderação de conteúdo vigente no Brasil, assim como sua percepção sobre pontos positivos, negativos e os impactos do Projeto de Lei 2.630/2020, voltado a regular plataformas no país.

Sobre o modelo regulatório de moderação de conteúdo vigente no Brasil

Neste tópico, os entrevistados foram questionados acerca da sua percepção em torno do modelo regulatório de moderação de conteúdo vigente no Brasil.

Para alguns deles, sequer haveria de se falar em um modelo brasileiro, isto é, **não haveria uma regulação específica voltada para a moderação de conteúdo no país (5x)**. O comentário de Fernanda Rezende, membro do Terceiro Setor, ilustra bem esse posicionamento:

Bom, eu teria a adicionar **que a gente não tem um modelo regulatório de moderação de conteúdo no Brasil**. Infelizmente, eu gostaria que a gente tivesse, mas a gente não tem. **A gente tem a autorregulação das empresas, que definem as suas regras**, e a gente tem, incipientemente, um arremedo disso com base na **relação institucional que tem, que está muito no início, do TSE com as plataformas**, mas um modelo regulatório no Brasil a gente não tem. (grifo nosso).

No ponto, destaca-se que as entrevistas foram realizadas em sua maioria no período pré-eleitoral de 2022, sendo que apenas uma delas foi realizada após. Em razão disso, a atuação mais incisiva do Tribunal Superior Eleitoral na moderação de conteúdo online, especialmente desinformativo, durante o pleito ainda não havia sido observada, senão somente as tentativas de cooperação entre o tribunal e as plataformas.

Acerca da inexistência de um modelo *sui generis* brasileiro, também foi mencionado que o Marco Civil da Internet (MCI) e a Lei Geral de Proteção de Dados (LGPD) seriam boas legislações, mas que estariam mais voltadas a uma regulação da internet, e não da moderação de conteúdo em si. Este último tema seria uma metaquestão da responsabilidade de intermediários, prevista no MCI, e seu debate mais específico **só teria sido iniciado com o Projeto de Lei 2.630/2020 (2x)**.

Para Augusto da Cruz, membro do setor governamental, as plataformas aplicariam o modelo americano em sua atuação no Brasil e seguiriam o MCI por ter princípios semelhantes ao primeiro, de forma que a regulação nacional seria fraca e necessitaria ser ampliada. Por outro lado, houve quem apontasse que o modelo brasileiro de moderação de conteúdo não seria inexistente, mas, sim, **insuficiente, tendo por base o Marco Civil da Internet (9x)**, principalmente, por meio do seu art. 19, que prevê a norma conhecida como *notice and take down*, ou seja, que estabelece a responsabilização de intermediários pelo conteúdo publicado por terceiros em suas plataformas somente quando não tomar as providências necessárias após a emissão de ordem judicial específica.

O Marco Civil seria uma lei, portanto, que **respondeu aos problemas de sua época (2x)**, mas sua **redação estaria defasada (3x)** e aquém da complexidade das demandas atuais, com **regras mínimas sobre responsabilização das plataformas (2x)** e que demandaria uma complementação e aprimoramento. Assim, na ausência de uma lei mais específica, criticou-se o **amplo poder conferido às empresas (2x)**, que poderiam abusar dele e remover conteúdos lícitos, sem direito a contraditório ou dever de explicação aos usuários, gerando insegurança jurídica. Ainda, a concessão de maior liberdade às plataformas não teria resultado, como esperado, em um combate mais eficiente a notícias falsas e discurso de ódio. O escrutínio pelo Poder Judiciário sobre sua atuação, por sua vez, também seria reduzido, por ser moroso e custoso, sendo que poucas pessoas poderiam acessar uma moderação por vias judiciais.

Em que pese as críticas, apontou-se que a criação de um modelo regulatório de moderação de conteúdo mais específico poderia ter por base o art. 19, do MCI, assim como o seu nível de abstração e caráter mais principiológico. Igor Peixoto, membro do setor empresarial, destacou bem a importância que o Marco Civil teve para o contexto de regulação da internet no Brasil:

Ele [o Marco Civil] **definiu a questão da responsabilidade de intermediários, o que foi super essencial naquele momento, que não tinha nem clareza exatamente, assim, quem que era o responsável pelo conteúdo postado nas redes sociais** e o regime, então, do art. 19 foi super importante em torno da ordem judicial, que foi ótimo, justamente **pra não estimular as empresas a remover conteúdo em excesso, que é o efeito colateral da pressão judicial, da pressão estatal**. E, ao mesmo tempo, é o sistema que deixou em aberto um **espaço para as plataformas definirem quais são as suas regras, para remover conteúdo com base nelas** (grifo nosso).

O espaço para as plataformas se desenvolverem também foi relacionado por Bárbara Silveira, do setor empresarial, que mencionou a possibilidade conferida a esses atores para que tomassem “medidas proativas sobre o conteúdo que pode ser legal, mas viola seus termos de serviço sem medo de litígios”.

Em virtude da sua relevância, outro argumento sugeriu colocar o MCI em revisão, a fim de verificar como se deu a sua aplicação e interpretação por tribunais e órgãos de aplicação, assim como a sua efetividade. Para Luiz Porto, do setor técnico-científico, essa proposta é mais interessante do que a criação de uma lei que pode vir a tirar o equilíbrio dos direitos e obrigações que aparecem no Marco Civil.

Ainda nesse tópico, os entrevistados também comentaram como avaliavam o papel dos atores envolvidos no modelo regulatório brasileiro, assim como se seriam necessárias mudanças para conciliar moderação de conteúdo e direitos humanos e quais seriam os mecanismos para garantir que os atores, incluindo intermediários, não subvertessem esse modelo de regime.

Para Samuel Cardoso, os três poderes estatais estariam com atuação aquém do esperado, seja por má informação, preocupação com pautas secundárias, como político-partidárias, ausência de especialização temática ou de atuação, dentre outros:

Pensando assim no aspecto do Estado, eu acho que como o **Poder Legislativo** está inerte, assim, eu não diria inerte, a gente tem tido muito debate, mas está mal informado, acho que o poder legislativo não trata com seriedade material a questão, acho que tem muito, **muita perspectiva política e pouca perspectiva técnica, tecnológica e pouca perspectiva técnica-jurídica e poucas preocupações efetivamente sociais nessa matéria e muitas preocupações político-partidárias, e muitas preocupações de demanda por poder e de atendimento a perspectivas limitadas**. Do ponto de vista do **Poder Judiciário**, acho que falta capacitação, falta atualização, faltam instâncias específicas para poder, em alguns casos, lidar com essas questões de forma mais especializada, **mas é o que tem feito alguma coisa, ainda que errando em várias situações, é o que tem sido, de certa forma, mais eficiente**. O **Poder Executivo** falha em realizar **políticas públicas**, em **promover debates** e a última grande medida de interferência em termos de moderação foi a medida provisória. Era um completo absurdo, era um erro gravíssimo que felizmente foi derrubado pelo funcionamento do controle de pesos e contrapesos da separação dos poderes no Brasil (...) (grifo nosso).

A crítica ao Legislativo vai ao encontro de outra percepção levantada, no sentido de haver receio quanto à atuação desse poder, que poderia resultar em uma má regulamentação. Segundo Vicente Moura, o poder econômico característico do setor privado nesse debate também seria capaz de influenciar no processo legislativo, e até mesmo na forma de interpretação do Poder Judiciário acerca de casos reais. Nessa senda, destacou a atuação de organizações da sociedade civil na “reação a eventuais jabutis que são inseridos, seja como proposta legislativa ou até mesmo novidades na plataforma, como alteração de políticas de privacidade ou termos de uso de diferentes plataformas, como foi o caso do Whatsapp”.

Entretanto, uma das percepções coletadas foi justamente no sentido de que o trabalho do Terceiro Setor nesse cenário tem sido voltado mais a uma postura reativa, em razão da quantidade de questões a serem enfrentadas e a falta de tempo hábil para tanto. Por outro lado, em relação à comunidade técnico-científica, argumentou-se haver um certo distanciamento pelo campo da tecnologia, que não estabeleceria um maior diálogo com os demais setores.

Já pensando em mecanismos para evitar a subversão do regime de moderação de conteúdo vigente no Brasil, seria necessária uma maior explicação e exposição das razões para a construção do patamar atual de moderação, que foi resultado de extensa análise baseada em experiências estrangeiras. Para Stephany das Neves, do setor técnico-científico, é preciso também um amadurecimento do próprio debate público em torno do que é moderação e o que se espera de uma lei que venha a regular a temática:

(...) eu acho que a gente precisa ter um debate público mais maduro, a gente precisa ter **clareza sobre quais são os pressupostos da moderação, quais são as funções da moderação, se ela é uma atividade legítima, em que medida, quais são seus limites**. Eu diria que a gente precisa de um debate público sério, amparado em fatos para que a gente possa de fato ter a **clareza de qual é a função da lei**. Esse talvez seja o principal problema hoje, a gente não tem clareza: **qual que é o objetivo da lei? Qual o objetivo da moderação e, portanto, qual o objetivo da lei?** (grifo nosso).

Para a entrevistada, apenas com essas respostas seria possível compreender se o regime está sendo subvertido ou não. Outro posicionamento também defendeu a necessidade de maior debate, porém entre os atores envolvidos no cenário regulatório, com atenção especial para as plataformas e para o que elas efetivamente seriam capazes de realizar em relação à aplicação de regras e tempo de solicitação. Da mesma forma, multas e sanções, tanto administrativas como aplicadas pelo Judiciário, poderiam atuar enquanto mecanismos para reforçar o cumprimento de dispositivos legais.

Para Lucas Oliveira, do Terceiro Setor, porém, já haveria um grande incentivo para que intermediários respeitassem regulações nesse sentido, uma vez que haveria uma predisposição das empresas em terceirizarem os problemas relacionados à moderação de conteúdo.

Sobre os aspectos positivos e negativos, assim como os impactos esperados a partir da aprovação do Projeto de Lei 2.630 (“PL das Fake News”)

Neste tópico, os entrevistados foram questionados acerca da sua avaliação em torno do Projeto de Lei 2.630, conhecido como PL das Fake News que, apesar do nome, tem na sua versão mais recente uma tentativa de regulação de plataformas sociais no Brasil. Conforme já mencionado, esse PL passou por intensas discussões, principalmente visando a sua aprovação antes das eleições de 2022, para que já pudesse ser aplicado durante esse período.

Nesse caminho, destaca-se novamente que a maioria das entrevistas foi realizada antes do pleito eleitoral, de tal modo que este foi o contexto de análise dos entrevistados sobre a proposta legislativa. Assim, eles destacaram, além de pontos positivos e negativos, os possíveis impactos que poderiam advir da aprovação do projeto.

Em relação aos aspectos positivos do PL 2.630, dois tópicos estiveram entre os mais citados: **garantia de maior transparência (11x)**, como através de obrigação de publicação de relatórios de transparência, e até mesmo em relação a anúncios; e a criação de **regras de um devido processo de moderação de conteúdo (6x)**, mas sem engessar a possibilidade de evolução desse sistema. Com relação à transparência, destacou-se, ainda, que ela poderia auxiliar no aumento da legitimidade do processo de moderação e da eficiência no direcionamento de pressão por parte da sociedade civil sobre essa pauta. Entretanto, haveria um certo exagero na quantidade de itens exigidos nos relatórios de transparência, a exemplo de dados específicos sobre as equipes de moderadores (sua configuração, nacionalidade, dados demográficos), o que poderia importar em riscos à privacidade.

A disponibilização de **mecanismos de contestação para o usuário (2x)** que não concorda com determinada remoção foi mencionado como um ponto positivo, pois também poderia auxiliar a trazer legitimidade e maior transparência para o procedimento. Da mesma forma, **racionalização de eventuais multas e sanções às plataformas (2x)**, ainda que em alguns casos pudesse ser vista alguma excessividade; a proibição de que agentes públicos, que usam as redes sociais de forma institucional, bloqueiem usuários; a disponibilização do termos de uso em linguagem acessível; avanços na pauta de criptografia em aplicativos de mensageria privada, com freios a eventual monitoramento de usuários; e a previsão de um espaço multissetorial para

auxiliar na definição de código de conduta de plataformas foram outros argumentos positivos levantados.

Com relação a esse último, porém, ele também foi apresentado como ponto negativo, pois uma entidade de autorregulação regulada seria interessante apenas para algumas plataformas, para outras não. Aliás, um dos problemas seria exatamente a busca por uma solução global que não enfrentasse as especificidades de cada plataforma.

Ainda assim, reconheceu-se que o PL 2.630 seria uma lei pragmática, que possuiria concessões para diferentes atores, como foi o caso da Lei Geral de Proteção de Dados e do Marco Civil da Internet. Nesse cenário, seria interessante **ter um regramento mínimo para as plataformas (3x)**, considerando que o vazio regulatório importaria em maior vulnerabilidade a todos e a proximidade com as eleições, pelo menos, à época das entrevistas.

Outrossim, o referido Projeto de Lei já estaria contribuindo com a **evolução do debate público brasileiro sobre moderação de conteúdo (2x)** e sobre a importância de que abusos na atuação de plataformas sejam denunciados. Além disso, ele buscaria melhorar a qualidade do conteúdo recebido pelos usuários e seria uma oportunidade para “aprofundar direitos de usuários da internet em redes sociais”, tendo por base o que já está previsto no Marco Civil da Internet.

Para Sofia Pires, membro do Terceiro Setor, a diferenciação feita para plataformas de grande escala também seria positiva ao ressaltar e proteger o restante da internet.

(...) e essa diferenciação mesmo que o projeto de lei faz entre plataformas de grande escala, então com número X de mais de 3 ou 4 milhões de usuários, alguma coisa assim, 10 milhões de usuários eu acho, e também a ponto de poder **ressalvar e continuar protegendo o resto da internet, o resto das plataformas que trabalham com conteúdo e não tem uma atuação muito incisiva ou que feche elas quase imediatamente**, assim, sabe? (grifo nosso).

Em sentido contrário, Samuel Cardoso apontou que a métrica de diferenciação de grandes plataformas, isto é, através do número de usuários, apresentaria desafios grandes de confiabilidade, como em razão de contas falsas e *bots*.

(...) ao mesmo tempo em que tem alguns critérios em relação aos quais eu tenho muita, muito ceticismo, por exemplo, de determinar que as plataformas tenham uma quantidade X de usuários. **A gente não tem uma métrica confiável para**

estabelecer quantos são os usuários do Twitter no Brasil, por exemplo - para dar o exemplo de uma rede relevante e que levanta dúvidas reais a respeito do seu tamanho no país ou no mundo como um todo. O quanto são bots, o quanto são perfis falsos, quantas pessoas distintas usam a mesma quantidade, a mesma conta ou enfim (grifo nosso).

Dentre os pontos negativos do PL, duas situações despontaram como as mais mencionadas: a **remuneração de veículos de notícias (12x)** e a previsão de **imunidade parlamentar em redes sociais (7x)**⁴¹. As razões para o primeiro ponto seriam: a) a complexidade da pauta, que demandaria um debate mais específico, e não apenas em um dispositivo de lei; b) o fato de que não teria relação com moderação de conteúdo para estar no PL; e c) seria um tema sensível, sem consolidação em outro lugar e sem previsões de como o procedimento ocorreria na prática. Para Guilherme Araújo, do setor privado, essa não seria uma pauta dentro da temática do PL 2.630:

A própria sociedade civil tem chamado atenção para a remuneração jornalística, **a gente precisa de debater isso em um outro ângulo, esse é um debate importante, acho que é um debate que toca mais... uma ou outra plataforma e não redes sociais ou mesmo aplicativos de mensageria**, esse é mais um tema relacionado aos mecanismos de busca, e esse é um dos graves problemas da peça não ser uma regulação tecnologicamente neutra, como é por exemplo a LGPD (grifo nosso).

Dentre os entrevistados, ainda houve crítica de que a proposta legislativa estaria mais focada neste tema, a partir da ideia de que maior investimento no jornalismo seria a solução para eliminar a desinformação, e que teria pouco resultado sobre a moderação de conteúdo. No entanto, uma ressalva foi feita no sentido de que monetizar o jornalismo poderia não ser de todo ruim, mas que não seria suficiente como solução para problemas envolvendo eventual descrédito dessa área, “no sentido de perder leitores”.

Quanto à imunidade parlamentar, foi argumentado que seria direito das pessoas verem o que autoridades públicas compartilham, sem limitação de acesso a esse

41 “(...) o parágrafo 8º do artigo [22], na atual versão do substitutivo [do Projeto de Lei 2.630/2020], estende a imunidade parlamentar material às redes sociais, impedindo que deputados e senadores sejam responsabilizados, civil e penalmente, por opiniões e palavras proferidas em ambiente digital. (...) Apesar do argumento de que o texto apenas reforça o que já está na Constituição Federal, sua presença em lei específica representa um alargamento da imunidade parlamentar, permitindo o entendimento de que esses atores estariam acima das regras de moderação de conteúdo de plataformas digitais”. COALIZÃO DIREITOS NA REDE. **Alerta de organizações da sociedade civil sobre o Projeto de Lei 2630/2020**. 6 de abril de 2022. Disponível em: <https://direitosnarede.org.br/2022/04/06/alerta-de-organizacoes-da-sociedade-civil-sobre-o-projeto-de-lei-2630-2020/>. Acesso em: 04 jan. 2022.

conteúdo. Para Lucas Oliveira, do Terceiro Setor, a previsão de uma imunidade parlamentar nas redes não estaria contemplada na Constituição Federal e também iria na contramão de pesquisas que demonstram o papel que atores políticos têm tido na propagação de desinformação online:

Então, por exemplo, aquela tentativa de colocar no PL uma imunidade parlamentar nas redes sociais. Pô, isso aí pra mim é impensável, é um tiro no pé sem... é um absurdo, né? Porque, assim, os parlamentos, obviamente, eles têm imunidade parlamentar, isso eu não posso nem discutir, isso é um dos pilares, inclusive, da nossa democracia, tá lá na Constituição, **só que aonde tá escrito que essa imunidade parlamentar se reflete também, também é refletida no ambiente digital, né?** E isso, inclusive, **contraria uma série de pesquisas empíricas que concluem que, hoje, os principais atores de desinformação na internet são aqueles que estão concorrendo em eleições, que ocupam cargos eletivos**, essas pessoas têm uma influência muito grande e o que elas postam nas plataformas acaba tendo um impacto, tanto é que quando o Twitter removeu o Donald Trump das plataformas, do Twitter especificamente, saiu uma pesquisa na sequência de um centro de estudos nos Estados Unidos concluindo que a desinformação sobre as eleições, sobre fraude eleitoral nos Estados Unidos caiu em 70% só com a remoção da conta do Trump, sabe? Então, assim, **focar nesses atores específicos é importante, inclusive, pra combater a desinformação nas plataformas digitais** (grifo nosso).

Outro ponto negativo levantado foi a **equiparação de redes sociais a meios de comunicação (2x)**, o que seria contrário ao que já está previsto no Marco Civil da Internet, que se preocupou em fazer a diferenciação de provedores de aplicação de internet que “não são atores dedicados a essa atividade-fim da comunicação”. A percepção de uma confusão trazida pelo PL 2.630 vai ao encontro de críticas apresentadas acerca da inserção de questões envolvendo o perfilamento de dados de usuários, quando já está vigente uma lei mais específica, como a LGPD, assim como o fato de que se trata de uma proposta que tenta abordar muitos assuntos ao mesmo tempo, sendo denominada como uma lei Frankenstein, ainda mais ao considerar que começou com o objetivo de combater desinformação e se tornou um PL de governança de plataformas.

Além disso, a falta de um debate público mais aprofundado sobre quais conteúdos devem ser moderados também impactaria na atuação do referido projeto. Para Stephany das Neves, membro do setor técnico-científico, a própria redação da proposta partiria de um pressuposto de que a moderação de conteúdo sempre viola

a liberdade de expressão, sendo necessário enxergar que essa atividade é legítima, para a partir de então estabelecer os seus limites. Segundo ela, não teria havido a criação de incentivos para que a moderação ocorresse “de uma forma adequada, reduzindo a desinformação, ampliando a pluralidade de debates”.

A esse respeito, Bárbara Silveira, do setor empresarial, destacou que, após alterações, o PL passou a tratar pouco sobre desinformação, de modo que seria mais interessante voltar a sua intenção original e combater este tema, em especial, no contexto político, “excluindo temas que nada contribuem ao enfrentamento da questão como o dispositivo mencionado acima” e trazendo “mais provisões que visem combater os incentivos financeiros dos agentes que criam e disseminam desinformação de maneira coordenada”. Se aprovado da forma como está, a entrevistada destacou que

(...) o texto que está para ser votado na Câmara dos Deputados irá **restringir o acesso das pessoas a fontes diversas e plurais de informação; desestimular as plataformas a tomar medidas para manter um ambiente saudável online; e causar um impacto negativo de negócios que buscam se conectar com seus consumidores por meio de anúncios e serviços digitais** (grifo nosso).

Ademais, **a criação de exigências e obrigações muito severas às plataformas (2x)** poderia resultar em prejuízos sociais e institucionais. O incentivo que o PL traria para a remoção de determinados conteúdos, por exemplo, poderia resultar em uma censura colateral. Dispositivos relacionados a anúncios e publicidade também foram mencionados, sugerindo-se que fossem mais enxutos, pois, atualmente, estariam servindo mais a interesses comerciais do que à sociedade.

Por fim, foram levantadas questões negativas que apareceram ainda na versão inicial do PL 2.630, no Senado Federal, como: criação de conceitos muito abertos, como era o caso de desinformação e contas inautênticas, por exemplo, o que novamente geraria estímulo para remoção em excesso; e questões envolvendo **rastreabilidade de mensagens instantâneas (4x)**, com ampliação da coleta e guarda de dados dos usuários; e o acesso a dados para fins de investigação.

Quando questionados, por sua vez, acerca de quais mecanismos seriam interessantes para garantir a efetividade de uma lei como o PL 2.630, no caso de sua aprovação, foi mencionado que o próprio procedimento de aprovação no legislativo seria um instrumento. Nesse sentido, o debate sobre a proposta e a construção de uma redação atenta à complexidade dos problemas que se apresentam deveria ser somado a uma ampla divulgação posterior da nova norma, a fim de explicar e capacitar institucionalmente tanto o Ministério Público quanto o Poder Judiciário, Executivo, escolas, comunidade acadêmica e científica.

Ainda, pontuou-se a possibilidade de criação de uma agência regulatória específica, voltada a acompanhar a atuação e relatórios de transparência publicados e a solicitar informações, mas considerando a necessidade de evitar a subversão a partir dos interesses políticos dos membros. O próprio Poder Judiciário, por meio da aplicação de medidas legais cabíveis em caso de descumprimento legal, também foi levantado como um mecanismo importante para **efetivar os dispositivos previstos no PL 2.630 (2x)**. A iniciativa da autorregulação regulada igualmente poderia auxiliar nesse cenário.

Para Emanuella Nogueira, entretanto, caberia ressaltar que determinadas obrigações previstas na proposta legislativa dependeriam não apenas das plataformas, mas também de outros atores, como é o caso de agentes políticos.

E também algumas das obrigações que estão ali não são questões que podem ser cobradas só das plataformas, mas também, por exemplo, de agentes públicos que utilizam as plataformas das redes sociais, a depender de algumas das regras ali têm consequências para outros atores, digamos assim, que mesmo do ponto de vista só da aplicação pelo Judiciário, eu **acho que podem ser bem efetivas dependendo de como sejam utilizadas, por atores políticos ou mesmo pela sociedade civil**, considerando que está lá para lidar com agentes políticos ou estatais e a forma como eles se utilizam dos canais de comunicação do poder público (grifo nosso).

Da mesma forma, uma instância de autorregulação regulada poderia vir a contribuir, mas seria necessário avaliar de que maneira ela se efetivaria na prática.

5. Discussão e considerações finais

O presente relatório teve por objetivo realizar o mapeamento das crenças, percepções e argumentos de especialistas acerca do papel desenvolvido pelos diferentes setores que atuam na regulação da moderação de conteúdo, assim como sobre os possíveis modelos de regime e a discussão em torno da construção de um modelo brasileiro. Os resultados obtidos demonstraram que, apesar de alguns consensos, ainda há questões que precisam ser melhor trabalhadas e mais aprofundadas no debate público, a fim de se chegar em uma regulação coerente com o contexto nacional.

Inicialmente, no que tange às concepções sobre as funções e ações que constituem as práticas de moderação de conteúdo, a ideia principal apresentada pelos entrevistados é de que moderar o espaço virtual constitui, basicamente, em deixá-lo “saúdável” e, assim, passível de uso e interação pelos usuários. Para isso, aplicar-se-ia ações

interventivas que apresentam como maior objetivo a remoção de conteúdos danosos, ilícitos e/ou ofensivos. Superado esse primeiro passo de caráter basilar, entende-se que a moderação contempla ações com objetivos além da simples sanitização, contemplando questões de curadoria de conteúdo e modulação de comportamentos. Esse ponto, por sua vez, se relaciona ao modelo de negócios das grandes plataformas, visto que, em prol de alcançar maior engajamento por parte dos usuários, os conteúdos são moderados em vista ao aumento do engajamento nas redes. Nesse sentido, aplicam-se ações como o ranqueamento de conteúdos, a rotulagem e os incentivos/desincentivos às postagens. Como consequência, observa-se uma maior polarização das redes, a qual se relaciona a fenômenos diversos, a exemplo dos filtros bolhas, atos de vigilância, extração de dados e modulações algorítmicas.

Em relação ao uso de mecanismos automatizados para a moderação de conteúdo, as discussões demonstraram o reconhecimento da importância do seu papel, ainda mais considerando a grande quantidade de informação disponibilizada na rede a cada segundo, bem como diante da sua eficiência para temas específicos, como pornografia infantil e direito autoral. Entretanto, a falha na análise de contexto ou linguagem e a possibilidade de existência de vieses nesses sistemas foram algumas das importantes ressalvas feitas em relação ao seu uso.

Com efeito, diferentes estudos já demonstraram que mecanismos de moderação automatizada podem apresentar erros ao analisar conteúdo publicado por grupos minoritários, por exemplo, drag queens⁴² e pessoas negras.⁴³ Assim, é necessário atenção para o risco de que esse tipo de ferramenta, ao invés de auxiliar a tornar o ambiente de plataformas digitais mais saudável, acabe silenciando pessoas que já são historicamente marginalizadas na sociedade.

Nesse sentido, alternativas como uma moderação por IA auditável, mais transparente, antidiscriminatória e atenta a contextos e especificidades regionais foram mencionadas pelos entrevistados, bem como a possibilidade de recurso de suas decisões e a necessidade de garantia de revisão humana. Em relação a esta última, apontou-se que deveria ser utilizada de maneira complementar à moderação automatizada, como uma espécie de filtro de casos sinalizados ou que passaram pela própria IA.

42 OLIVA, Thiago Dias; ANTONIALLI, Dennys Marcelo; GOMES, Alessandra. Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture*. v. 25. n. 2, p. 700–732 abr. 2021. Disponível em: <https://link.springer.com/article/10.1007/s12119-020-09790-w> Acesso em: 20 dez.. 2022.

43 HAIMSON, Oliver L.; DELMONACO, Daniel; NIE, Peipei; WEGNER, Andrea. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. In: ACM CONFERENCE ON HUMAN-COMPUTER INTERACTION, 2021. *Proceedings [...]* Association for Computing Machinery, 2021. Disponível em: <https://dl.acm.org/doi/10.1145/3479610>. Acesso em: 19 dez. 2022.

Entretanto, foi possível constatar uma importante preocupação com o psicológico de moderadores humanos, que seriam expostos a grandes quantidades de conteúdo nocivo em um curto espaço de tempo - tema este que também vem sendo explorado na literatura.⁴⁴ A delegação desse tipo de trabalho a países em desenvolvimento e a informação reduzida sobre times de moderação foram mencionados como agravantes desse cenário.

Em face à responsabilidade dos agentes diante das práticas de moderação de conteúdo, foi possível observar que o setor privado, representado pelas grandes plataformas e conglomerados de tecnologia, é compreendido como o setor que detém os maiores poderes e, logo, deveres no campo. Nesse cenário, a ausência de diretrizes regulatórias expressas e marcos legais específicos para as ações de moderação de conteúdo gera uma situação ambígua, em que não existem obrigações determinadas, mas também não há limites expressos. Logo, pela situação de privilégio que as plataformas possuem em questão de disponibilidade de recursos e pela natureza de suas atividades, atribui-se a elas a competência de moderar.

Embora o setor privado tenha sido eleito como o maior responsável frente às atividades de moderação, os entrevistados se apresentaram majoritariamente céticos em relação aos recursos disponíveis para a fiscalização da sociedade diante de tais práticas. Nesse cenário, volta à tona a questão da ausência de diretrizes expressas para conduzir a moderação de conteúdo. Ao passo que não há regras específicas para a publicação dos relatórios de transparência, por exemplo, não haveria motivo pelo qual as grandes empresas de fato se comprometeriam com os princípios de transparência e accountability. Essa questão, por sua vez, se relaciona também à priorização do modelo de negócios das plataformas, uma vez que práticas de transparência absoluta poderiam acarretar em prejuízos comerciais ou no cumprimento de condições não tão interessantes ao setor privado no que diz respeito aos interesses econômicos em vista.

Nesse contexto, insere-se a discussão quanto às instâncias privadas de revisão de decisão de moderação de conteúdo, sendo o Oversight Board o principal exemplo considerado. O caráter inédito e ainda noviço dessa iniciativa foi um ponto levado em consideração, levando a ressalvas para opinar sobre e/ou a posições em valorização à iniciativa e expectativas quanto ao seu amadurecimento. Assim, observou-se que, apesar da maioria dos entrevistados defender a atuação positiva do órgão, diversas críticas foram apresentadas. O foco principal de tais ponderações foi quanto à atuação e autonomia limitada das instâncias privadas, visto que partem de uma iniciativa

44 STEIGER, Miriah et al. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In: CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2021, Yokohama. **Proceedings [...]** Yokohama: Association for Computing Machinery, 2021. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3411764.3445092>. Acesso em: 20 dez. 2022.

das próprias plataformas e, assim, teriam sua imparcialidade comprometida. Em relação ao OB, especificamente, observou-se que o funcionamento do órgão suscita dúvidas e críticas. Visto que o órgão se propõe a analisar poucos casos para, a partir deles, publicar recomendações de caráter sistêmico, as principais críticas foram endereçadas justamente ao fato de não serem avaliados mais casos - além dos paradigmáticos, - de maneira a gerar uma atuação de “maior impacto”. Ademais, as críticas ao impacto foram também atreladas à opacidade do órgão, visto a ausência de parâmetros apontada pelos entrevistados para a condução das decisões.

Quanto à possibilidade de atuação do Poder Judiciário enquanto possível instância de revisão judicial de decisões de moderação, alguns entrevistados pontuaram que o próprio Marco Civil da Internet havia lhe capacitado enquanto parte essencial dessa atividade. Isto, pois ao prever, em seu art. 19, que os intermediários só poderiam ser responsabilizados pelo conteúdo publicado por terceiros em suas plataformas no caso de descumprimento de ordem judicial específica, empoderou-se os tribunais como uma importante figura nesse ecossistema. É de se observar, porém, que tal argumento conflita com outro apresentado, no sentido de que fugiria à competência do Judiciário atuar como instância revisora dessa atividade.

Por outro lado, além de não haver como escapar da possibilidade de buscar o Judiciário em razão da garantia de acesso à justiça a todo cidadão, ele também seria fundamental para atuar na solução de conflitos ou restrição de direitos, da mesma forma que na correção de ilegalidades e reparação de danos eventuais causados pela atividade de moderação. No entanto, levar uma quantidade muito grande de demandas desse tipo para a via judicial poderia significar contribuir para a sobrecarga que já é observada nas cortes. Ainda, o descompasso entre a velocidade de evolução da internet e do Judiciário, bem como a falta de varas especializadas na temática dificultariam uma atuação mais assertiva dessa instituição, sendo que haveria também um risco de concessão de decisões inadequadas.

A respeito da atuação do terceiro setor e academia nos processos regulatórios de conteúdo online, as posições apresentadas pelos entrevistados foram majoritariamente defendendo seu caráter positivo; no entanto, as críticas estiveram presentes na quase totalidade das falas. Assim, conclui-se que o ponto mais recorrente foi que tais setores apresentam contribuições significativas à qualificação do debate público, com destaque para a participação em processos legislativos e exercendo pressão diante das práticas realizadas pelo setor privado e governamental. No entanto, encontram limitações diversas em suas atividades, principalmente em face à escassez de recursos e, logo, maiores dificuldades para obter influência significativa no campo.

No que diz respeito à criação de regras pelas próprias plataformas, observa-se um entendimento majoritário em concordância à necessidade de que elas tenham essa

autonomia - visto que é preciso a existência de regras internas para a determinação de condutas aceitas e, logo, responsabilidades às próprias plataformas. No entanto, houve críticas à forma como essas regras são desenvolvidas, com foco para o caráter opaco de seus parâmetros e diretrizes (ou mesmo inexistência) e para as dificuldades de implementação e aplicação, as quais, por não apresentarem parâmetros bem estabelecidos, podem ocorrer de forma seletiva e imparcial.

Nesse cenário, outra questão pontuada foi quanto à dificuldade em desenvolver regras, a partir da autorregulação, que são de fato conduzidas em atenção aos princípios de Direitos Humanos. Foi ressaltado que traçar esses parâmetros de forma universal é uma atividade delicada, que envolve a consideração a questões de diversidade - ainda mais quando as plataformas possuem uma atuação global. No entanto, cabe às plataformas a atenção a essas questões no momento de construção e aplicação de suas regras, o que grande parte dos entrevistados apontou como algo que não acontece.

Quanto à regulação estatal, observou-se que a maioria dos entrevistados entende a aplicação de normas pelos agentes estatais como necessárias no campo da moderação de conteúdo. Essa posição, por sua vez, dialoga com questões já levantadas, tal como a frustração diante da ausência de diretrizes regulatórias específicas no que tange à moderação de conteúdo, inexistência ou ineficiência de mecanismos de fiscalização quanto à atuação do setor privado e posição de privilégio por parte das plataformas na gestão da moderação de conteúdo. Assim, foi apontado que a regulação estatal adotaria o papel principal de estabelecer padrões mínimos no campo, ausentes hoje em dia - criando assim obrigações de transparência, por exemplo. No entanto, críticas também foram apresentadas em vista às possibilidades de abuso, censura e frente às dificuldades de regular a área de tecnologia, dada a forte possibilidade de obsolescência das legislações diante da velocidade dos processos e transformações.

No que tange ao regime intitulado autorregulação regulada, foi possível verificar um certo consenso de que a sua definição seria a possibilidade de as próprias plataformas estabelecerem suas regras, mas com base em diretrizes previstas pelo Estado e sem que o regulador estabelecesse o conteúdo dessa autorregulação. Todavia, houve discussões no sentido de que esse regime, na verdade, seria uma espécie de correção e logo, na possibilidade de demandar ou não a existência de uma instância de controle.

Ainda assim, ele foi apontado como um modelo que poderia ser mais eficiente do que os outros, além de ser a melhor opção no momento, seguindo uma tendência verificada em outras iniciativas nacionais. Este posicionamento também pode ser verificado na literatura especializada, como em Hartmann e Sarlet, que defendem ser a autorregulação regulada a melhor alternativa atualmente, sendo “aceitável que as plataformas possuam alguma autonomia na gestão da manifestação”, mas sem descartar um controle externo delimitado em suas características e condições.⁴⁵

A forma como esse regime foi introduzido no texto do Projeto de Lei 2.630/2020, chamado “PL das Fake News”, foi alvo de duras críticas, como por permitir a interpretação de que a criação de uma órgão de revisão de decisões de moderação fosse condição de existência a esse regime. Além disso, criticou-se a estipulação do Comitê Gestor da Internet como instância responsável por dialogar com as plataformas acerca da elaboração de códigos de conduta. De modo mais geral, aspectos negativos sobre a autorregulação regulada também disseram respeito ao receio de abuso econômico por parte das plataformas e de que a atuação de um órgão regulador pudesse não funcionar adequadamente no cenário brasileiro.

No que diz respeito aos mecanismos de Soft Law diante da adequação das práticas de moderação de conteúdo aos direitos humanos, tem-se que a ausência de força legal e a liberdade das plataformas em poder definir como será de fato sua aplicação foram questões criticadas pelos entrevistados. Desse modo, o tema encontra elementos correspondentes em tópicos anteriores, tal como na responsabilidade da gestão da moderação de conteúdo e no ceticismo dos entrevistados para com a implementação de mecanismos de fiscalização. Observa-se assim que, mais uma vez, o caráter discricionário das plataformas diante da definição das regras de moderação de conteúdo e a ausência de diretrizes expressas advindas de mecanismos legais faz com que apenas elementos em acordo aos modelos de negócios sejam considerados. No entanto, cabe mencionar que muitos entrevistados citaram pontos positivos nas recomendações sem efeito legal, a exemplo da condução de parâmetros para a criação de regras pelas plataformas e marcos regulatórios, execução de pressão para com as plataformas e criação de espaços de debate.

No caso de criação de normas estatais estabelecendo diretrizes procedimentais para a moderação de conteúdo, houve ampla manifestação positiva, com destaque para a elaboração de bases mínimas para um procedimento ideal e de um devido processo para essa atividade, estabelecendo garantias processuais mínimas aos usuários. Dentre elas, mencionou-se: quais as hipóteses e o que será feito nas atividades de moderação, motivos e regras utilizadas para uma decisão, possibilidade de recurso, prazos para revisão de determinados conteúdos e regras para notificação de usuários que tiveram conteúdo moderado.

Outros exemplos de diretrizes também foram apresentados e merecem destaque: mecanismos a serem observados na remoção de conteúdo; canais específicos para oferecimento de explicações sobre remoções; políticas de governança de conteúdo em linguagem acessível; elaboração de relatórios de transparência e o tipo de informação que devem conter; parâmetros de tomada de decisão e critérios para moderação; incentivos a ações alternativas à remoção, como sinalização ou

diminuição do alcance de conteúdo e checagem de fatos com rede de especialistas contratados; informações acerca do uso de inteligência artificial; avaliação do alinhamento do procedimento a parâmetros internacionais de direitos humanos e do desempenho das plataformas na aplicação dos seus termos de serviço; e respeito a direitos fundamentais, como a liberdade de expressão.

Essas conclusões foram fundamentais, pois demonstram que disposições previstas no PL 2.630/2020 sobre devido processo estão de acordo com o posicionamento de especialistas da área. Entretanto, vale mencionar que ressalvas como a necessidade de se evitar um engessamento do procedimento de moderação e de respeito às particularidades de diferentes plataformas estão no radar para delimitar o alcance desse tipo de norma.

Ainda acerca de legislações estatais, os argumentos dos entrevistados se apresentaram bem divididos quando se tratava de regras voltadas à determinação de remoção de conteúdos específicos. Com a edição da Medida Provisória 1.068/2021, pelo governo Bolsonaro, um sinal de alerta foi acendido para a possibilidade de que normas nesse sentido fossem utilizadas de maneira equivocada. Dentre aqueles que se posicionaram a favor desse tipo de lei, destaca-se a menção a conteúdos específicos que poderiam ser objeto, como racismo, homofobia e transfobia. Entretanto, houve preocupação com a ausência de definições mais específicas em torno de conceitos considerados “zonas cinzas”, como seria o caso da liberdade de expressão, e de uma compreensão uniforme do Poder Judiciário sobre temas como discurso de ódio.

O foco em conteúdos já considerados ilegais no ordenamento jurídico vigente, isto é, exceções previstas em lei ou condutas já consideradas criminosas, seria uma das principais alternativas para evitar possíveis consequências negativas. Nesse turno, merece destaque o receio de que leis determinando a remoção de conteúdo específicos pudessem gerar mais incentivo à censura (tanto por parte do Estado, como pelas próprias plataformas) e a uma remoção excessiva pelos intermediários, na busca por evitar a responsabilização.

Em relação à regulação da moderação de conteúdo no Brasil, foi possível verificar uma necessidade geral de avanço em previsões específicas para essa área. As opiniões dos entrevistados dividiram-se, principalmente, no sentido de que ou o modelo vigente brasileiro seria inexistente ou insuficiente, sendo que nesse caso apenas o já mencionado art. 19, do Marco Civil da Internet, seria a norma mais palpável a esse respeito. Entretanto, a redação do MCI estaria defasada e aquém da complexidade de questões em evidência atualmente.

Ainda assim, elaborar uma regulamentação que tivesse por base o Marco Civil, observando-se o seu nível de abstração e caráter principiológico, ou colocá-lo em revisão, diante de seus anos de vigência, foram colocações que ressaltaram

a importância dessa lei para o contexto da regulação da internet no Brasil. Essas questões são fundamentalmente importantes diante do debate promovido pelo PL 2.630/2020, que enfrentou estipulações específicas para a moderação de conteúdo.

Nesse cenário, novamente a questão referente à falta de uma regulação voltada para o campo foi mencionada, argumentando-se que essa ausência abriria margem para que as empresas tivessem amplo poder sobre a moderação, o que facilitaria abusos, como a remoção de conteúdos lícitos sem direito a contraditório ou maiores explicações aos usuários. Em virtude dos custos e morosidade, contar com a avaliação do Poder Judiciário sobre a sua atuação seria uma possibilidade reduzida, pois poucas pessoas poderiam acessar a via judicial.

De forma mais direcionada, acerca da avaliação de aspectos positivos e negativos do Projeto de Lei 2.630, alguns assuntos ganharam maior relevância na discussão. Em torno de pontos favoráveis, os tópicos mais mencionados foram a garantia de maior transparência e a criação de regras de um devido processo de moderação através da proposta. Na sequência, a disponibilização de mecanismos de contestação para o usuário; racionalização de eventuais multas e sanções às plataformas; a proibição de que agentes públicos, que usam as redes sociais de forma institucional, bloqueiem usuários; a disponibilização do termos de uso em linguagem acessível; avanços na pauta de criptografia em aplicativos de mensageria privada, com freios a eventual monitoramento de usuários; e a previsão de um espaço multissetorial para auxiliar na definição de código de conduta de plataformas foram outros pontos positivos levantados.

É interessante notar que alguns desses pontos já haviam sido avaliados como positivos em perguntas anteriores não direcionadas ao PL 2.630, como é o caso da criação de um devido processo e o estabelecimento de mecanismos como relatórios de transparência. Nessa direção, o projeto vai ao encontro de quem já havia se manifestado favoravelmente a normas estatais que preveem diretrizes procedimentais mínimas para a moderação de conteúdo.

Por outro lado, no que tange a aspectos negativos, o destaque foi para as previsões de remuneração de veículos de notícias, que deveria ser pauta de um regramento específico para esse fim, e imunidade parlamentar em redes sociais, que iria na contramão de pesquisas que apontam o papel de atores políticos na propagação de desinformação online. Somado a eles, também foram mencionados: a equiparação de redes sociais a meios de comunicação; falta de um debate público aprofundado sobre quais conteúdos deveriam ser moderados; redução de previsões em torno de desinformação; criação de exigências e obrigações muito severas às plataformas, que poderiam resultar em aumento de censura, dentre outros.

Esses exemplos demonstram que, apesar do avanço trazido nas discussões em torno do PL 2.630, muitos pontos negativos ainda poderiam ser melhorados para uma versão final. Passada a pressa para buscar a sua aprovação antes do pleito eleitoral de 2022, os próximos anos possibilitarão que o debate em torno da regulação de plataformas, assim como da moderação de conteúdo, seja mais aprofundado e estudado entre as diferentes partes interessadas. Dessa forma, ter em mente estes e outros pontos de melhoria apresentados no presente trabalho pode ser um bom ponto de partida para que eventuais equívocos não sejam repetidos em uma nova tentativa de legislação.

De todo modo, o que ficou demonstrado do estudo realizado é que persistem marcados dissensos quanto a diversos aspectos da governança de conteúdo online. Diversos interesses estão em jogo, tanto econômicos quanto sociais e políticos, além da disputa por proteger direitos fundamentais básicos, como é o caso da liberdade de expressão e de manifestação. O local cada vez maior que grandes plataformas sociais vêm ganhando na vida das pessoas, porém, não deixa dúvidas de que esse debate é necessário, a fim de que abusos não sejam cometidos e para que usuários possam se sentir seguros nos espaços virtuais.

Nesse sentido, muitas críticas foram pontuadas, demonstrando insatisfação diante da assimetria de poderes que compõem o cenário da Governança de conteúdo online hoje, principalmente no Brasil, em que ainda não há uma regulação específica a delimitar o amplo campo de atuação das plataformas. No entanto, também não foram propostas soluções que pretendem ser rápidas e instantâneas, mesmo porque há o entendimento entre os atores de que, para resolver as principais questões, é preciso equilibrar interesses, em respeito ao princípio do multissetorialismo.

Entre governo, sociedade civil, comunidade técnico-científica e empresariado, a regulação da moderação de conteúdo exige um diálogo aberto, transparente e realista, que possa garantir a criação de um arcabouço normativo que faça sentido no contexto e diante das demandas específicas do território brasileiro.

6. Referências bibliográficas

ACCESS NOW. **26 recommendations of content governance**: a guide for lawmakers, regulators, and company policy makers. Disponível em: <https://www.accessnow.org/cms/assets/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf>. Acesso em: 19 dez. 2022.

ARCHEGAS, João Victor; GODOY, Miguel Gualano de. Os limites da jurisdição do Facebook Oversight Board: de Marbury v Madison para Facebook v Trump. **Jota Info**, 02 de fevereiro de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/os-limites-da-jurisdicao-do-facebook-oversight-board-02022021>. Acesso em: 20 dez. 2022.

COALIZÃO DIREITOS NA REDE. **Nota técnica sobre o relatório de 26 de junho de 2020 ao Projeto Lei nº 2.630/2020**. 28 de junho de 2020. Disponível em: <https://direitosnarede.org.br/2020/06/28/nota-tecnica-sobre-o-relatorio-de-26-de-junho-de-2020-ao-projeto-lei-no-2-630-2020/>. Acesso em: 19 dez. 2022.

COMITÊ DE SUPERVISÃO. **Comitê de Supervisão**. 2022. Disponível em: <https://oversightboard.com/>. Acesso em: 30 ago. 2022.

DENARDIS, Laura. **The Global War for Internet Governance**. Yale University Press, 2014.

DOURADO, Tatiana Maria Silva Galvão. **Fake news na eleição presidencial de 2018 no Brasil**. 308 f. Tese (Doutorado) – Programa de Pós-Graduação em Comunicação e Culturas Contemporâneas, Universidade Federal da Bahia, Salvador, 2020.

EFF – Electronic Frontier Foundation et al. **Princípios de Manila sobre responsabilidade civil de intermediários**. Disponível em: <https://manilaprinciples.org/pt-br/principles.html>. Acesso 20 dez. 2022.

EFF – Electronic Frontier Foundation et al. **Santa Clara Principles on transparency and accountability in content moderation**. Disponível em: <https://santaclaraprinciples.org/pt/cfp/>. Acesso 20 dez. 2022.

ESTARQUE, Marina; ARCHEGAS, João Victor. **Redes sociais e moderação de conteúdo**: criando regras para o debate público a partir da esfera privada, Instituto de Tecnologia e Sociedade (ITS): Rio de Janeiro, 2021, p. 20. Disponível em: <https://itsrio.org/pt/publicacoes/redes-sociais-e-moderacao-de-conteudo/>. Acesso em: 19 dez. 2022.

GAMA, Sophia. Guerra de desinformação: as fake news nas eleições de 2018. **Câmara**

Municipal de Curitiba, 15 de julho 2022. Disponível em: <https://www.curitiba.pr.leg.br/informacao/noticias/guerra-de-desinformacao-as-fake-news-nas-eleicoes-de-2018>. Acesso em: 19 dez. 2022.

GORWA, Robert. The platform governance triangle: conceptualising the informal regulation of online content. **Internet Policy Review**, v. 8, n. 2, jun./2019. Disponível em: <https://policyreview.info/articles/analysis/platform-governance-triangle-conceptualising-informal-regulation-online-content>. Acesso em 30 ago. 2022.

GORWA, Robert. What is platform governance? **Information, Communication & Society**, v. 22, n. 6, pp 854-871, 2019. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1573914>. Acesso em: 24 ago. 2022.

HAIMSON, Oliver L.; DELMONACO, Daniel; NIE, Peipei; WEGNER, Andrea. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. In: ACM CONFERENCE ON HUMAN-COMPUTER INTERACTION, 2021. **Anais [...]** Association for Computing Machinery, 2021. Disponível em: <https://dl.acm.org/doi/10.1145/3479610>. Acesso em: 19 dez. 2022.

HARTMANN, Ivar; IUNES, Julia. Fake news no contexto de pandemia e emergência social: os deveres e responsabilidades das plataformas de redes sociais na moderação de conteúdo online entre a teoria e as proposições legislativas. **RDP**, Brasília, v. 17, n. 94, p. 388-414, jul./ago. 2020. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/download/4607/Hartmann%3B%20Iunes%2C%202020>. Acesso em 19 dez. 2022.

HARTMANN, Ivar Alberto Martins; SARLET, Ingo Wolfgang. Direitos fundamentais e direito privado: a proteção da liberdade de expressão nas mídias sociais. **Revista de Direito Público**, v. 16, n. 90, p. 85-108, dez. 2019. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3755>. Acesso em: 19 dez. 2022.

HOFMANN, Jeanette. Multi-stakeholderism in Internet governance: putting a fiction into practice. **Journal of Cyber Policy**, v.1, n. 1, pp. 29-49. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/23738871.2016.1158303>. Acesso em: 11 jan. 2023.

KADRI, Thomas E.; KLONICK, Kate. Facebook V. sullivan: Public figures and newsworthiness in online speech. **Southern California Law Review**, v. 93, n. 37, pp. 37-99, 2019, p. 94. Disponível em: https://scholarship.law.stjohns.edu/cgi/viewcontent.cgi?article=1292&context=faculty_publications. Acesso em: 30 ago. 2022.

KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. **Harvard Law Review**, v. 131, n. 6. p. 1598-1669, abr. 2018. Disponível em: <https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/>. Acesso em: 30 ago. 2022.

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: fev, 2022.

MARS, Amanda. Como a desinformação influenciou nas eleições presidenciais? **El País**, 25 de fevereiro 2018. Disponível em: https://brasil.elpais.com/brasil/2018/02/24/internacional/1519484655_450950.html. Acesso em: 19 dez. 2022.

OLIVA, Thiago Dias; ANTONIALLI, Dennys Marcelo; GOMES, Alessandra. Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. **Sexuality & Culture**. v. 25. n. 2, p. 700–732 abr. 2021. Disponível em: <https://link.springer.com/article/10.1007/s12119-020-09790-w>. Acesso em: 20 dez. 2022.

RODRIGUES, Gustavo Ramos. “A função de uma lei de proteção de dados é proteger a todos, inclusive aquele que coleta dados”: o debate sobre a Lei Geral de Proteção de Dados e a ideologia da harmonia no Fórum da Internet no Brasil. In: VII ENCONTRO NACIONAL DE ANTROPOLOGIA DO DIREITO, 2022, São Paulo. **Anais [...]**. São Paulo: NADIR/USP, 2021. v. 1. p. 12. Disponível em: <https://bit.ly/3GBIZYx>. Acesso em: 11 jan. 2023.

SILVA JUNIOR, L. A.; LEAO, M. B. C. O software Atlas.ti como recurso para a análise de conteúdo: analisando a robótica no Ensino de Ciências em teses brasileiras. **Ciência & Educação**, Bauru, v. 24, n. 3, p. 715-728, set. 2018. Disponível em: <https://www.scielo.br/j/ciedu/a/yBwC9L74v4vD3s4PwVXggsk/?lang=pt>. Acesso em: 21 dez. 2022.

STEIGER, Miriah et al. The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support. In: CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS, 2021, Yokohama. **Anais [...]** Yokohama: Association for Computing Machinery, 2021. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3411764.3445092>. Acesso em: 20 dez. 2022.

SARLET, Ingo Wolfgang; SIQUEIRA, Andressa de Bittencourt. Liberdade de expressão e seus limites numa democracia: o caso das assim chamadas “faew news” nas redes sociais em período eleitoral no Brasil. **Revista Estudos Institucionais**, v. 6, n. 2, pp.



534-578, maio/ago. 2020. Disponível em: <https://www.estudosinstitucionais.com/REI/article/view/522/511>. Acesso em: 19 dez. 2022.

TWITTER. **Conselho de Trust & Safety**. 2022. Disponível em: <https://about.twitter.com/pt/our-priorities/healthy-conversations/trust-and-safety-council>. Acesso em 30 ago. 2022.

Apêndice I - Roteiro da entrevista

Bloco I - Apresentação e temas gerais

1. Qual sua área de formação e principal atividade?
2. Você já atuou, na pesquisa, na prática, na regulação, em relação a moderação de conteúdo?

Explorar: em que atividades relacionadas a moderação de conteúdo online, seja na prática, na regulação ou na pesquisa, você já atuou?

Bloco II - Moderação de conteúdo e a interação entre os diferentes atores envolvidos

3. No seu entendimento, quais as principais funções da moderação de conteúdo?
Explorar: quais atividades você considera que são moderação de conteúdo? Por quê?

4. Entre todos os atores da governança da internet, existe algum ator com maior responsabilidade sobre a gestão de moderação de conteúdo?

Explorar: por quê? Quais são as responsabilidades? Como os outros atores podem cooperar com elas?

5. Na sua percepção, existem mecanismos para garantir a fiscalização da sociedade sobre práticas de moderação de conteúdo?

Explorar: quais seriam eles? São eficazes?

6. Hoje em dia, existem instâncias privadas de revisão de decisões de moderação de conteúdo, como o Oversight Board do Facebook, por exemplo. De modo geral, como você avalia a atuação dessas instâncias?

Explorar: o que tem de positivo? E de negativo? Como você avalia que isso impacta na legitimidade do procedimento? E qual o papel do Poder Judiciário enquanto possível instância de revisão judicial da moderação de conteúdo?

7. Como você avalia a atuação dos sujeitos do terceiro setor e comunidades técnico científicas nos processos regulatórios de conteúdo online?



Explorar: você avalia como positiva ou negativa? Em que situações a sociedade civil é mais mobilizada no contexto da Governança da internet? Quais organizações são mobilizadas?

Caso a avaliação seja negativa: você pensa que há uma marginalização? Você vê alguma solução para isso?

Caso seja interessante: como você avalia o debate público sobre moderação de conteúdo em plataformas sociais hoje?

Bloco III - Modelos regulatórios

8. Pensando na adequação da moderação a direitos humanos, como você enxerga a efetividade da criação de regras de conteúdo pelas próprias plataformas?
9. Existem também as recomendações sem efeito legal, como o Princípios de Santa Clara, por exemplo, que estabelecem padrões de transparência para práticas de moderação de conteúdo. Ainda pensando na adequação a direitos humanos, como você avalia o impacto desse tipo de padrões e recomendações sem efeito legal?
10. Como você avalia o papel das normas emitidas por autoridades estatais para a adequação da moderação de conteúdo aos direitos humanos?
11. Em uma escala de 0 a 10, o quanto você é favorável à criação de leis que obriguem plataformas a removerem determinado tipo de conteúdo por conta própria? Por quê?

Explorar: se não conseguir dar uma nota: quais dificuldades você visualiza em dar uma nota nesse caso? Se a resposta for intermediária: por que não 0/10? Se depende do caso, quais são eles? Quais tipos de conteúdo? Seria possível conciliar isso com garantia de direitos dos usuários?

12. Você considera que a lei deveria estabelecer diretrizes específicas para os procedimentos de moderação de conteúdo?

Explorar: Que mecanismos poderiam garantir a efetividade dessa lei? O que essas diretrizes deveriam assegurar? Como considerar plataformas que contam com moderação voluntária ou de menor escala?

13. Como você avalia o uso de mecanismos automáticos para remoção ou restrição de conteúdo? Por quê?

Explorar: há riscos específicos associados a eles? Se sim, quais? Como você avalia a revisão humana em relação a esses mecanismos?

14. Você já ouviu a expressão “autorregulação regulada”?

Se sim: quais as fontes? O que entende por isso? Como você avalia essa proposta? Na sua avaliação, ela é mais ou menos efetiva do que as outras?

Se não: existem autores que mencionam esse termo, a partir de obras alemãs, como uma forma de estabelecer padrões para a regulação interna dos procedimentos de moderação de conteúdo; como você avalia essa proposta? Na sua avaliação, ela é mais ou menos efetiva do que as outras?

15. Na sua percepção, como você enxerga o modelo regulatório de moderação de conteúdo vigente no Brasil hoje?

Explorar: como você avalia o papel dos atores envolvidos no modelo regulatório brasileiro? Na sua percepção, seriam necessárias mudanças para conciliar moderação de conteúdo e direitos humanos? Quais?

Se for interessante dependendo da pessoa: quais seriam os mecanismos para garantir que os atores, incluindo intermediários, não subvertam esse modelo de regime?

16. Você está acompanhando o debate sobre o PL 2630 (“PL das Fake News”)?

Explorar: no caso de implementação do PL 2630, em trâmite no Legislativo brasileiro, como você avalia que seriam os efeitos sobre a moderação de conteúdo e os direitos dos usuários no Brasil? Quais são os mecanismos necessários para a efetividade dessa regulação? Quais são os pontos positivos e negativos do PL?

Alternativa, se a pessoa não souber do PL: quais são os mecanismos necessários para garantir a efetividade da regulação sobre transparência na moderação de conteúdo? Quais seriam os pontos positivos e negativos desses mecanismos?

17. Você quer adicionar algum comentário?

18. Tem alguma pergunta que não foi feita e você considera que seria importante?



Apêndice II - Listagem de códigos utilizados

1. S. Academia
1. S. Governamental
1. S. Privado
- 1.S. Terceiro setor
2. MOD ação ::
2. MOD ação :: curadoria (modular comportamento/ranqueamento de conteúdo)
2. MOD ação :: intervenção (remoção/redução/exclusão de conteúdo)
2. MOD função
2. MOD função :: ambiente saudável (sanitização do espaço virtual/ambiente receptível)
2. MOD função :: produto
3. R1 - plataformas :: recurso (expertise/poder econômico)
3. R2 - Estado :: poder jurisdicional (fiscalização/regulação/legislação)
3. Responsável pela gestão - outro
3. Responsável pela gestão – governo
3. Responsável pela gestão – plataforma
- 4 MF - inexistente ::
4. MF - insuficiente ::
4. MF - suficiente ::
5. IP - legitimidade não
5. IP - legitimidade não determinada

- 5. IP - legitimidade sim
- 5. IP - negativo :: autonomia limitada
- 5. IP - positivo :: outro
- 5. IP – negativo ::
- 5. IP – negativo :: falta de transparência
- 5. IP – positivo ::
- 5. IP – positivo :: transparência
- 5. PJ - colisão de direitos
- 5. PJ - negativo ::
- 5. PJ - negativo :: competência
- 5. PJ - negativo :: escala
- 5. PJ - neutro
- 5. PJ - positivo
- 6. 3º setor e academia - atuação negativa
- 6. 3S e academia - intermediação
- 6. 3S e academia - limit. ::
- 6. 3S e academia - limit. :: autonomia limitada (recursos/financiamento)
- 6. 3S e academia - mob ::
- 6. 3S e academia - mob :: episódios focais
- 6. 3S e academia - particularidades (impacto nas plataformas/pressão)
- 6. 3S e academia - qualificação do debate público
- 6. 3S e academia – proatividade



- 7. AR - dificuldades
- 7. AR - negativo
- 7. AR - positivo :: necessidade
- 8. RE - negativo :: censura
- 8. RE - negativo :: competência
- 8. RE - positivo
- 9. SL - negativo
- 9. SL - negativo :: limitações
- 9. SL - positivo
- 9. SL - positivo :: impacto/pressão
- 10. Diretrizes procedimentais - negativo
- 10. Diretrizes procedimentais - positivo
- 11. Conteúdo regulável - negativo
- 11. Conteúdo regulável - Nota
- 11. Conteúdo regulável - positivo
- 12. IA - negativo
- 12. IA - negativo :: erro contextual
- 12. IA - positivo
- 12. IA - positivo :: necessidade
- 13. Revisão humana - positivo:: revisar erros
- 13. Revisão humana - negativo
- 13. Revisão humana - positivo

14. AR-R - concepção

14. AR-R - negativo

14. AR-R - positivo

15. BR atual - inexistente

15. BR atual - infos relevantes

15. BR atual - insuficiente

15. BR atual - negativo

15. BR atual - positivo

15. BR atual - suficiente

16. PL 2630 - impactos

16. PL 2630 - negativo

16. PL 2630 - negativo :: imunidade parlamentar

16. PL 2630 - negativo :: remuneração imprensa

16. PL 2630 - positivo

16. PL 2630 - positivo :: devido processo

16. PL 2630 - positivo :: transparência



INSTITUTO
DE REFERÊNCIA
EM INTERNET
E SOCIEDADE