

Nota Técnica sobre Procedimentos de Moderação de Conteúdo (art. 12 do PL 2630/2020)

I. Justificativa

A moderação de conteúdo, entendida como o conjunto de mecanismos de governança que estruturam a participação em uma plataforma para facilitar cooperação e prevenir abusos,¹ vem sendo largamente empregado pelos provedores de redes sociais e de aplicações que permitam compartilhamento de conteúdo na internet, bem como nos resultados de buscas.²

Apenas no primeiro quarto de 2020, o Facebook removeu ou aplicou medidas sobre cerca de 1,8 bilhões de contas inautênticas e conteúdos envolvendo nudez e atividade sexual, violência, organizações perigosas (terrorismo e ódio), *hate speech*, drogas e armas de fogo, pornografia infantil, suicídio e auto-mutilação. O último Relatório de transparência do Twitter revela suspensão de 244.188 contas por exploração de menores (91% usando tecnologia) e 115.861 contas por conteúdo violento (87% com uso de ferramentas computacionais proprietárias).³ Segundo o Relatório do primeiro trimestre de 2020,⁴ moderadores do Youtube removeram manualmente 399.422 vídeos, ao passo que 5.711.586 vídeos foram removidos por ferramentas automatizadas.⁵ O relatório de transparência do Google, por sua vez, indica que mais de 4 bilhões de URLs tiveram sua deslistagem requisitada por usuários.

Pela posição central das plataformas digitais na construção do espaço de debate público atual, a tendência, na Europa⁶ e nos Estados Unidos da América,⁷ consiste em ver a moderação como central não só para a promoção da qualidade dos serviços disponibilizados, que traduz o interesse pri-

¹ GRIMMELMANN, James. The virtues of moderation. *Yale Journal of Law & Technology*, 2015, v. 17. Disponível em: <<https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?article=1110&context=yjolt>>. Acesso: em 27 ago.2020.

² Barrett, Paul. M. Who moderates the social media giants? In: *New York University Stern Center for Business and Human Rights*, 2020. Disponível em: <<https://bhr.stern.nyu.edu/blogs/2020/6/4/who-moderates-the-social-media-giants>>. Acesso em: 27 ago.2020.

³ TWITTER. **Transparency. Rules Enforcement.** Disponível em: <<https://transparency.twitter.com/en/twitter-rules-enforcement.html>>. Acesso em: 28 ago.2020.

⁴ GOOGLE **Transparency Report.** Cumprimento das diretrizes da comunidade do You Tube. Disponível em: <<https://transparencyreport.google.com/youtube-policy/removals>>. Acesso em: 28 ago.2020.

⁵ GOOGLE **Transparency Report.** Cumprimento das diretrizes da comunidade do You Tube. Disponível em: <https://transparencyreport.google.com/youtube-policy/removals?total_removed_videos=period:Y2020Q1;exclude_automated:human_only&lu=total_removed_videos>. Acesso em: 28 ago.2020.

⁶ JØRGENSEN, Rikke Frank. **When Private Actors Govern Human Rights. Research Handbook on Human Rights and Digital Technology** 346, 363. In: Ben Wagner, Matthias C. Kettemann, Kilian Vieth (Orgs.) 2019; ROBERTS, Sarah T. Behind the Screen. **Content moderation in the shadows of social media**, Edward Elgar: Londres, 2019 p. 33 ss.

Barrett, Paul, **Regulating Social Media: The Fight Over Section 230 — and Beyond**, September 2020. Senator Hawley Introduces Bill to Remove Section 230 Immunity from Behavioral Advertisers, Office of Senator Josh Hawley, July 28, 2020, <https://www.hawley.senate.gov/senator-hawley-introduces-bill-remove-section-230-immunity-behavioral-advertisers>, CITRON, Danielle K; FRANKS, Mary Anne. The internet as a speech machine and other myths confounding section 230 Reform. **Boston University School of Law Public Law and Legal Theory Paper**, 2020, n. 20-8; ⁷ CITRON, Danielle K; WITTES, Benjamin. The internet will not break: Denying bad Samaritans § 230 immunity. **Fordham Law Review**, v. 86, pp. 401–423.

vado dos provedores, mas também para a garantia de liberdade de expressão do indivíduo, de direitos fundamentais e do ambiente democrático, razão pela qual, há uma necessidade de convergência dos termos de uso com a ordem jurídica Estatal, em democracias constitucionais contemporâneas. Por outro lado, cada vez mais se reconhece que essa curadoria de conteúdo implica responsabilidade das plataformas pela adoção dos procedimentos adequados de governança e meios técnicos disponíveis para evitar abusos e mitigar impactos provocados por conteúdo nocivo propagado. Ou seja, aponta-se para deveres inerentes à moderação de conteúdo pelas plataformas terem se tornado, de fato, as anfitriãs do espaço para o discurso público e “curadoras” da liberdade de expressão, de modo que suas regras não podem ficar restritas à relação contratual com os usuários, devendo incorporar a legislação nacional, sendo compatíveis com as regras sobre licitude de conteúdo difundido na esfera pública.⁸ Isso inclui não só os ilícitos criminais,⁹ como também as regras nacionais sobre publicidade¹⁰ nos veículos de comunicação, nas práticas de propaganda online e impulsionamento de conteúdo.

Tendo em vista o interesse público do qual se cerca, os sistemas privados de governança montados pelos provedores para a moderação de conteúdo na internet tem sido centro de debates internacionais, que apontam preocupação quanto à falta de transparência e arbítrio pelos provedores.¹¹ Nesse sentido, entidades da sociedade civil tem promovido iniciativas e promulgado documentos no sentido de estimular a adoção de mecanismos de transparência, dentre eles o *Santa Clara Principles*¹², o *Corporate Accountability Index*¹³ e os princípios sobre filtragem de conteúdo da *Electronic Frontier Foundation*¹⁴.

Esses documentos elencam a necessidade de provedores de aplicação divulgarem os números de medidas aplicadas (*flagging, blocking, takedown*) e de notificarem o usuário afetado pela medida, oferecendo as razões correspondentes, bem como a oportunidade para sua contestação. Iniciativas legislativas também incorporaram alguns deveres procedimentais da transparência, como a lei ale-

⁸ WIELSCH, Dan. Os ordenamentos das redes: Termos e condições de uso - Código - Padrões da comunicação. em: Ricardo Campos, Georges Abboud, Nelson Nery Jr. (Orgs.) **Fake News e Regulação**, 2a. Edição, RT: Sao Paulo 2020, p. 91 ss.

⁹ No Brasil, no mínimo, as práticas capituladas nos tipos penais presentes nos artigos 138,139,140, 147, 282, 283 e 284 da Lei 7.209 de 11 de julho de 1984 (Código Penal), na Lei 7.716 de 5 de janeiro de 1989 e artigos 323 e 337 do Código Eleitoral e artigos 33, parágrafo 4o e 34, parágrafo 3o da Lei 9.504 de 30 de setembro de 1997.

¹⁰ No Brasil, naquilo que for cabível, as Leis nº 4.680/65 e a Lei nº 12.232/10 e ao disposto no art. 36 da Lei 8.078/90.

¹¹ HAMILTON, Rebecca J. Governing the Global Public Square. **Harvard International Law Journal**, v. 62, 2021 (no prelo); JØRGENSEN, Rikke Frank; ZULETA, Lumi. Private Governance of Freedom of Expression on Social Media Platforms. **Nordicom Review**, 2020, pp. 51-67; CAPLAN, Ronyn; GILLESPIE, Tarleton. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. **Social Media + Society**, 2020; GORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. **Big Data & Society**, 2020; KETTEMANN, Matthias. Menschenrechte und politische Teilhabe im digitalen Zeitalter, 2020; TIEDEKE, Anna Sophia; KETTEMANN, Matthias. Back up: can users sue platforms to reinstate deleted content? **Internet Policy Review Journal on internet regulation**, vol. 9, 2020.

¹² THE SANTA CLARA PRINCIPLES on Transparency and Accountability in Content Moderation. Disponível em: <<https://santaclaraprinciples.org/>>. Acesso em: 26.ago 2020.

¹³ Ranking Digital Rights Header Branding. Disponível em: <<https://rankingdigitalrights.org/index2019/>>. Acesso em: 26.ago 2020.

¹⁴ GEBHART, Gennie. Who has your Back? Censorship Edition 2019. **Electronic Frontier Foundation**. Disponível em: <<https://www.eff.org/wp/who-has-your-back-2019>>. Acesso em: 26.ago 2020.

mã de redes *NetzDG*¹⁵ e a proposta de regulação europeia sobre prevenção de disseminação online de conteúdo terrorista¹⁶ e a Recomendação sobre a Responsabilidade de Intermediários na Internet do Conselho Europeu.¹⁷

Embora as plataformas tenham passado a divulgar números de atividade de moderação, induzidas por pressões governamentais,¹⁸ foram tímidas em relação à disponibilização de mecanismos efetivos de contestação, permanecendo arbitrárias as decisões pelos provedores quanto à exclusão de conteúdos e contas.¹⁹

Por outro lado, o debate acadêmico e de ativistas digitais²⁰ tem apontado a insuficiência dessas medidas. Note-se que todas elas se referem a boas práticas *após a tomada de decisão* pelo provedor de aplicação acerca de medida sobre conteúdos ou contas. Mas uma série de preocupações têm sido levantadas sobre a adequação e impactos a direitos fundamentais decorrentes do próprio processo decisório adotado pelas plataformas. Esse processo de tomada de decisão envolve, basicamente, equipes de moderadores e ferramentas computacionais para a detecção ou decisão automatizada sobre o conteúdo.

Quanto aos moderadores, há três ordens de questionamento. A primeira diz respeito à adequação dos padrões de moderação, tipicamente modulados para democracias constitucionais, assumindo-se, por exemplo, que a desinformação e o ódio são criados e tolerados apenas por uma minoria, quando na verdade podem ter o suporte da maioria da população e ter inclusive origem governamental.²¹ A segunda concerne à necessidade de competência dos moderadores para lidar com questões locais, incluindo o conhecimento da língua e aspectos culturais e políticos relevantes, considerando que, por vezes, a moderação não é nacional e até mesmo pode ser executada de modo terceirizado para empresas estrangeiras.²² Recentemente, o *Facebook* fez esforço louvável em cons-

¹⁵ *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*, Seção 2, que requer a produção de relatórios de transparência. Ver tradução da lei alemã para o português em: Ricardo Campos, Georges Abboud, Nelson Nery Jr. (Orgs.) **Fake News e Regulação**, 2a. Edição, RT: Sao Paulo 2020, p. 337 ss.

¹⁶ EUROPEAN COMMISSION. Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, 2018. O Art. 8(2) do referido regulamento obriga a publicação de requerimentos de remoção de conteúdo.

¹⁷ https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680790e14

¹⁸ MASNICK, Mike. How Government Pressure Has Turned Transparency Reports from Free Speech Celebrations to Censorship Celebrations, *Techdirt*, publicado em 17.abr 2018. Disponível em: <<https://www.techdirt.com/blog/?d=17&m=4&y=2018>>. Acesso em: 26.ago 2020.

¹⁹ GILLESPIE, Tarleton. Custodians of the Internet. Platforms, content moderation, and the hidden decisions that shape social media, **New Haven** 2018, p. 24 ss.; KLONICK, Kate. The New Governors: The People, Rules, and Processes Governing Online Speech” 131. **Harvard Law Review**, 1598.

²⁰ Ver globalvoices.org e datasociety.net.

²¹ WU, Tim. Is the First Amendment Obsolete? **Michigan Law Review**, vol. 117, 2018, pp. 547-557; HAMILTON, Rebecca J. Governing the Global Public Square. **Harvard International Law Journal**, v. 62, 2021 (no prelo);

²² GILLESPIE, Tarleton. Custodians of the Internet. Platforms, content moderation, and the hidden decisions that shape social media, New Haven. Yale University Press, 2018; ROBERTS, Sarah T. Digital detritus: ‘Error’ and the logic of opacity in social media content moderation. *First Monday*, 23(3); SUZOR, Nicolas P.; WEST, Sarah Myers; QUODLING, Andrew; YORK, Jillian. What do We mean when We talk about transparency? Toward meaningful transparency in commercial content moderation. **International Journal of Communication** 2019, vol 13. Disponível em: <<https://ijoc.org/index.php/ijoc/article/view/9736>>. Acesso em: 26.ago 2020.

tituir e divulgar seu *Oversight Board*²³ para moderação de conteúdo, enfatizando a equipe escolhida, formada por membros conhecidos por sua defesa da liberdade de expressão e direitos humanos. Porém, o Oversight Board decide pequeno número casos selecionados, com o objetivo de oferecer diretrizes para as equipes de moderadores, que efetivamente tomarão decisões sobre moderação de conteúdo na rede social. Pouco se sabe sobre a composição, treinamento e capacidades das equipes de moderadores, em qualquer das redes sociais hoje dominantes. O episódio de Myanmar,²⁴ em que o *Facebook* foi incapaz de detectar como a rede social foi utilizada como arma para incitar violência contra a minoria étnica *Rohingya*, ilustra bem essas duas ordens de preocupação. A terceira preocupação concerne às condições de trabalho do moderador, que é exposto constantemente a imagens e conteúdo degradantes, o que pode afetar sua saúde psíquica.²⁵

Por sua vez, a automação na moderação para a detecção e tomada de decisão de conteúdo abusivo, tanto *ex ante* quanto *ex post*, tem sido estimulada pela regulação como o *Copyright Directive* da União Européia, o *NetzDG* alemão e a lei australiana *Abhorrent Violent Material*, que obrigam a adoção de medidas em tempo eficiente, por vezes dentro de 24 horas. Porém, críticos dessa legislação apontam para o risco de excesso de confiança nessas ferramentas²⁶ e que a discricionariedade das plataformas no seu uso pode levar a resultados contraproducentes,²⁷ como violações à liberdade de expressão, pelo excesso de bloqueios,²⁸ e a outros direitos humanos, caso não haja transparência e prestação de contas em relação a essas ferramentas,²⁹ em particular quanto a seu papel na tomada de decisão, quanto a sua explicabilidade, i.e. critérios relevantes para a classificação, muitas

²³ Disponível em: <<https://www.oversightboard.com/news/announcing-the-first-members-of-the-oversight-board/?fbclid=IwAR1o-ZWZjr3eJqa06kwLkWJZTaSG7H8M9bnPz2qNu0RlbBNSXopEjPnZwZul>>. Acesso em: 27.ago 2020.

²⁴ STECKLOW, Steve. Inside Facebook's Myanmar Operation Hatebook: A Reuter's Special Report, REUTERS, publicado em 15.ago 2018. Disponível em: < <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/> [<https://perma.cc/2QWF-D6L8>]. Acesso em: 26.ago 2020.

²⁵ ROBERTS, Sarah T. Behind the Screen. **Content moderation in the shadows of social media**, Yale University Press: Londres, 2019 p. 33 ss.

²⁶ LI, Sydney; WILLIAMS, Jamie. Despite What Zuckerberg's Testimony May Imply, AI Cannot Save Us. **Electronic Frontier Foundation**, publicado em 11.abr 2018.. Disponível em: <<https://www.eff.org/deeplinks/2018/04/despite-what-zuckerbergstestimony-may-imply-ai-cannot-save-us>>. Acesso em: 26.ago 2020.

²⁷ BLOCH-WHEBA, Hannah. Automation in Moderation, **Cornell Internattional Law Journal** 2020.

²⁸ BAR-ZIV, Sharon; ELKIN-KOREN, Niva. Behind the scenes ofonline copyright enforcement: Empirical evidence on notice & take-down. **Connecticut Law Review**, 2018; URBAN, Jennifer; KARAGANIS, Joe; SCHOFIELD, Brianna. Notice and takedown in everyday practice. **UC Berkeley Public Law Research Paper**, 15.mar 2016.

²⁹ GORWA, Robert; BINNS, Reuben; KATZENBACH, Christian. Algorithmic content moderation: Technical and political challenges in the automation of platform governance. **Big Data & Society** 2020.

vezes tornada opaca pelo algoritmo empregado,³⁰ seu nível de acurácia e a possível incorporação de traços sociais discriminatórios³¹ ou distorções em acurácia conforme gênero ou raça.³²

Por fim, considerando que, embora não respondam por sua elaboração, a propagação dos conteúdos é viabilizada pela plataforma, de modo que tem sido destacada a necessidade dos provedores de aplicação adotarem os meios técnicos a seu alcance visando a mitigar impactos eventualmente provocados pelas ações ou omissões da moderação. Tal atuação inclui não só restaurar contas e conteúdos equivocadamente excluídos, como também empregar meios disponíveis para corrigir a informação ou informar sobre o caráter abusivo do conteúdo àqueles que a ele foram expostos, ou no mínimo, que com ele interagiram.³³

A partir do debate internacional recente, é possível observar que o tratamento dado pelo PL nº 2630/2020 à moderação de conteúdo é incompleto. Isso porque, além de não estipular a necessidade dos termos de serviço dos provedores de aplicação incorporarem parâmetros de conteúdo ilícito, conforme legislação nacional sobre a comunicação e sobre a publicidade, tratam apenas de um aspecto da moderação, a saber, aquilo que ocorre após a tomada de decisão dos provedores sobre o conteúdo, com a obrigatoriedade de oferecer oportunidade de defesa e recurso, além da divulgação de métricas da moderação.

Faltam regras de transparência e prestação de contas naquilo que ocorre no processo de tomada de decisão pela plataforma, seja em relação aos moderadores, seja em relação aos algoritmos empregados, bem no que diz respeito à disponibilização de mecanismos claros para reclamação para os usuários ofendidos e ao empenho dos provedores no sentido de mitigar impactos do conteúdo nocivo que foi propagado na rede.

Nesse sentido, os seguintes pontos merecem especial atenção no tema da moderação de conteúdo privado pelos provedores:

1. As regras elaboradas pelo provedor de aplicação sobre moderação de conteúdo devem ser debatidas de modo amplo, à luz de seus impactos específicos sobre o discurso público, envolvendo diferentes representantes da sociedade civil, considerando-se o papel relevante do

³⁰ ANANNY, Mike.; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 2018, 20(3), pp. 973–989. GORWA, Robert; Garton ASH, Timothy Garton. Democratic Transparency in the Platform Society. In: Persily N, Tucker Josh, **Social Media and Democracy: The State of the Field and Prospects for Reform**. Cambridge, UK: Cambridge University Press, 2020.

³¹ BINNS, Reuben, VEALE, Michael, KLEEK, Max Van; SHADBOLT, Nigel. Like trainer, like bot? Inheritance of bias in algorithmic content moderation. In: **International conference on social informatics**, pp. 405–415. Berlin: Springer, 2017.

³² ANGWIN, Julia; GRASSEGGGER, Hannes. Facebook's Secret Censorship Rules Protect White Men, But Not Black Children. In: **ProPublica**. Disponível em: <[org/article/facebook-hate-speech-censorship-internaldocuments-algorithms](https://www.propublica.org/article/facebook-hate-speech-censorship-internaldocuments-algorithms)>. Acesso em: 25.ago 2020; HOFFMANN, Anna Lauren. Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. **Information, Communication & Society** 2019, vol. 22, ed. 7, pp. 900–915.

³³ ECKER, Ullrich K H; SWIRE-THOMPSON, Briony. Misinformation and its Correction: Cognitive Mechanisms and Recommendations for Mass Communication (2018). Disponível em: https://www.researchgate.net/publication/317603082_Misinformation_and_its_Correction_Cognitive_Mechanisms_and_Recommendations_for_Mass_Communication

discurso veiculado nas plataformas para o exercício de direitos individuais, para a coesão social e para a democracia.³⁴ Tanto as regras como as ferramentas para sua aplicação, incluindo a moderação automática e humana devem ser descritas de modo transparente. As plataformas devem responder tanto pelas regras quanto por sua aplicação, assegurando que adota os procedimentos e medidas adequadas, tendo em vista seu papel essencial como fóruns do discurso público e como palcos para exercício de uma gama de direitos humanos online, incluindo liberdade de expressão, mas também a liberdade de informação, direito à dignidade e à privacidade.

2. A moderação de conteúdo *online* contra a desinformação deve abranger não somente serviços de redes sociais, que permitam ao usuário publicar conteúdos *online* para redes de relacionamentos ou ao público em geral (*Facebook, Twitter, etc.*), como também serviços de postagem e compartilhamento de vídeos (*You Tube*), bem como serviços de busca que direcionem o usuário para conteúdo desinformativo (*Google search*). Assim o artigo sobre moderação não deve limitar sua referência a provedores de redes sociais, mas abranger os provedores de aplicação em geral;
3. Atenção deve ser dada, também, à moderação de publicidade. O provedor tem a oportunidade e o dever de filtrar o conteúdo, previamente, ao seu impulsionamento, considerando que obterá retorno financeiro pela difusão. Esse dever precisa estar aliado às regras de transparência para identificação das fontes de financiamento de propagação de conteúdos ilícitos, submetendo-se à regulação existente sobre publicidade no Brasil;
4. A atividade de moderação humana deve contar com equipe de revisores adequada ao volume de comunicações, com competência linguística e conhecimento da cultura local adequados para o exame do conteúdo;
5. Ferramentas computacionais que empreguem inteligência artificial cada vez mais tem se mostrado eficazes e efetivas em custos para detetar e classificar conteúdo ilícito. Porém, são falíveis e devem atender a parâmetros éticos para evitar que sejam contraproducentes ou ameacem outros direitos fundamentais, dentre eles, padrões aceitáveis de acurácia para seu emprego, transparência, explicabilidade e controle de vieses;
6. Dado o caráter falível da moderação e a possibilidade de exclusão ou marcação da expressão pública de conteúdo lícito e legítimo, o poder de moderação das plataformas não pode ser arbitrário, devendo incorporar também mecanismos online de contestação, facilmente acessíveis pelo usuário e que garantam o efetivo contraditório, além de respostas rápidas e justificadas pelos provedores quanto a sua decisão final;
7. Deve haver transparência pelos provedores em relação à equipe empregada para moderação de conteúdo, empresas contratadas e aos meios técnicos, em particular, sobre os sistemas de inteligência artificial utilizados;
8. Os provedores de aplicação devem, na medida do possível naquilo que for tecnicamente viável, buscar informar usuário que teve contato com conteúdo que violou os termos e usos

³⁴ Cf. KETTEMANN/SCHULZ, [Setting Rules for 2.7 Billion](#). A (First) Look into Facebook's Norm-Making System (Leibniz Institute for Media Research Working Paper #2, 2020).

do contrato ou que alvo de medidas de checagem por fontes independentes. Essa medida informativa pela plataforma vem em consonância com seus deveres de tráfego visando a contribuir com a atenuação dos efeitos da circulação e propagação de desinformação ou conteúdo abusivo. Essa medida deve atingir, no mínimo, aqueles que visualizaram ou interagiram explicitamente com o conteúdo abusivo pelo período em que foi mantido na rede e quando possível, todos aqueles que foram expostos ao conteúdo. Provedores de aplicação, em suas especificidades técnicas, utilizam métricas diversas para mensurar engajamento e monitoramento dos usuários que efetivamente tiveram contato com os diferentes tipos de conteúdos disponibilizados. Nesse contexto, é importante que o texto consiga abarcar a pluralidade de métricas para garantir a eficácia da norma e o alcance da medida para o máximo de usuários afetados, conforme a viabilidade técnica³⁵.

³⁵ Sobre as métricas utilizadas veja Facebook: <https://www.facebook.com/business/help/743427195703387>, Youtube: <https://support.google.com/youtube/answer/2991785?hl=pt-BR>, Twitter: <https://help.twitter.com/pt/using-twitter/media-studio-analytics>, Métricas que consideram engajamento: <https://www.facebook.com/business/help/251837856080253?helpref=search&sr=1&query=engagement>

II. Proposta de redação para o art. 12 do PL 2630/2020

Art. 12. Os procedimentos de moderação de conteúdo consistem nos mecanismos de governança que estruturam a participação em plataforma de internet para prevenir abusos e violações aos termos de uso dos provedores de aplicação e compreendem medidas para a classificação, tomada de decisão sobre o conteúdo postado ou contas em operação, devendo abranger:

I- regras adequadas que observem e promovam direitos fundamentais e o Estado de Direito no ambiente online;

II- equipe adequada de moderadores e ferramentas computacionais confiáveis;

III- canal online efetivo para recebimento e apuração de denúncias por usuários;

IV- procedimento em plataforma online para recurso contra decisões pelos provedores de aplicação que garanta o devido processo; e

V- meios técnicos para mitigar o impacto de conteúdo ilícito propagado na plataforma.

§ 1º Os termos de uso dos provedores de aplicação deverão incluir, dentre os conteúdos e as contas em operação sujeitos à aplicação de medidas, aqueles que veiculem discurso ou práticas manifestamente ilegais, que constituam desinformação, calúnia, difamação, injúria, ameaça e incitação à violência, exercício ilegal da medicina, arte dentária ou farmacêutica, charlatanismo, curandeirismo, crimes resultantes de preconceito de raça ou de cor, exploração sexual de menores ou crimes eleitorais, nos termos da legislação vigente.

§ 2º Os termos de uso em relação à publicidade online e impulsionamento de conteúdo deverão observar as normas de publicidade no País, em especial as Leis nº 4.680/65 e n. 12.232/10, além da Lei nº 8.078/90, devendo os provedores de serviço de aplicação distinguir claramente perante o usuário o conteúdo patrocinado e publicidade, bem como moderar seu conteúdo previamente à difusão publicitária ou impulsionamento.

§ 3º A equipe de moderadores deverá ter dimensão proporcional ao volume de usuários da plataforma, receber treinamento adequado para a função e acesso a tratamento profissional para preservação de sua saúde psíquica, possuir competência linguística e conhecimento da cultura local para o exame do conteúdo.

§ 4º A aplicação de medidas sobre conteúdo e contas em operação com base em decisões automatizadas deve ser transparente quanto ao papel da automatização, bem como propiciar informações gerais sobre os seus critérios de operação, padrões de acurácia e medidas para controle de vieses.

§ 5º Deverá ser disponibilizado aos usuários canal facilmente acessível para denúncias sobre conteúdo e contas em operação.

§ 6º Os provedores de aplicação de internet submetidos a esta Lei devem assegurar, nos processos de elaboração e aplicação de medidas com base em seus termos de uso, mecanismos de recurso e devido processo.

§ 7º Em caso de denúncia ou de medida aplicada em função dos termos de uso das aplicações ou da presente Lei que recaia sobre conteúdos e contas em operação, o usuário deve ser notificado sobre a

fundamentação, o processo de análise e a aplicação da medida, assim como sobre os prazos e procedimentos para sua contestação.

§ 8º Os provedores dispensarão a notificação aos usuários se verificarem risco:

I – de violação a direitos de crianças e adolescentes;

II – de crimes tipificados na Lei nº 7.716, de 5 de janeiro de 1989;

III – de grave comprometimento da usabilidade, integridade ou estabilidade da aplicação.

IV – de postagem reiterada de conteúdo idêntico ou similar a conteúdo sobre o qual já foi adotada medida pelo próprio provedor ou por outros provedores de aplicação.

§ 9º Deve ser garantido pelo provedor o direito do usuário recorrer de decisões adotadas pelo provedor de aplicação sobre conteúdos e contas, bem como corrigir classificações equivocadas de conteúdo, em tempo útil e eficaz, com a restauração de contas indevidamente suspensas ou excluídas por violações aos termos de uso.

§ 10º O prazo de defesa será diferido nos casos de conteúdo que use imagem ou voz manipuladas para imitar a realidade, com o objetivo de induzir a erro acerca da identidade de candidato a cargo público, ressalvados o ânimo humorístico ou de paródia.

§ 11º A decisão do procedimento de moderação deverá assegurar ao ofendido o direito de resposta na mesma medida e alcance do conteúdo considerado abusivo.

§ 12º O provedor de aplicação deverá notificar os usuários acerca de adoção de medida sobre conteúdo com o qual tenha havido interação ou tenha sido visualizado, por meio de qualquer ferramenta disponível na plataforma, fornecendo, quando possível, informação acurada sobre o tema, proveniente de fonte independente.

§ 13º Os relatórios de transparência, previstos no § 1º do art. 13, deverão informar sobre os procedimentos e resultados da moderação, incluindo informações sobre empresas eventualmente contratadas para moderação, quantidade de pessoas envolvidas, bem como sobre métricas de publicidade e conteúdo impulsionado.

III. Equipe do Instituto LGPD responsável pelo documento



Juliano Maranhão. Diretor do Instituto LGPD. Bacharel, doutor e livre docente pela Faculdade de Direito da Universidade de São Paulo, onde atualmente é professor associado do Departamento de Filosofia e Teoria Geral do Direito. Pesquisador da Fundação Alexander von Humboldt e professor convidado da Goethe-Universität Frankfurt am Main. Foi pesquisador visitante nas universidades de Miami, Leipzig e Maastricht. Pós-doutorado no departamento de ciência da computação da Universidade de Utrecht. Membro do Comitê Executivo da *International Association for Artificial Intelligence and Law* (IAAIL) e do Conselho Editorial do *Artificial Intelligence and Law Journal*.



Ricardo Campos. Diretor do Instituto LGPD. Docente assistente (wissenschaftlicher Mitarbeiter) na cátedra de direito público, direito das novas mídias e teoria do direito na faculdade de Direito da Goethe Universität Frankfurt am Main, Alemanha. Mestre em Teoria do Direito (LL.M – Master of Laws) pela Goethe Universität, Frankfurt am Main (2010). cursou graduação em Direito (2008) na Universidade Federal de Juiz de Fora, na Goethe Universität Frankfurt am Main (Alemanha) e na Universität Passau (Alemanha).



Juliana Abrusio. Diretora do Instituto LGPD. Doutora em Direito pela Pontifícia Universidade Católica de São Paulo. Mestre em Direito pela Università degli Studi di Roma Tor Vergata, com diploma revalidado pela Universidade de São Paulo (USP). Professora da Faculdade de Direito da Universidade Presbiteriana Mackenzie. Professora Convidada da Fundação Dom Cabral. Professora convidada da Pós Graduação da POLI – Universidade de São Paulo. Professora convidada do programa de Mestrado Profissional do IDP-SP. Coordenadora do Comitê de Direito Digital do CESA – Centro de Estudos das Sociedades de Advogados. Vice Presidente da Comissão de Direito Digital do IASP – Instituto dos Advogados de São Paulo.



Matthias Kettemann. Consultor do instituto LGPD. Mestre pela Universidade de Harvard, Doutor pela Universidade de Graz (Austria) e livre-docente pela Goethe Universität em Frankfurt (Alemanha). Dirige o programa de governança na Internet do renomado Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI) em Hamburg (Alemanha). Foi consultor em questões digitais das principais instituições e empresas na Alemanha e Europa dentre elas parlamento alemão (Bundestag), partidos políticos, várias empresas indexadas na bolsa de valores alemã, fundações e organizações internacionais, incluindo o Conselho da Europa, o Parlamento Europeu e a OSCE sobre regulamentação da Internet, segurança cibernética e direitos humanos.



Giovanni Sartor. Consultor do instituto LGPD. Professor de informática jurídica na Universidade de Bolonha e professor de Informática Legal e Teoria Jurídica no Instituto Universitário Europeu de Florença. Doutor pelo Instituto Universitário Europeu (Florença), trabalhou no Tribunal de Justiça da União Europeia (Luxemburgo), foi pesquisador do Conselho Nacional Italiano de Pesquisa (IT-TIG, Florença), ocupou a cadeira de Direito na Universidade Queens of Belfast (onde agora é professor honorário), e foi professor Marie-Curie na Universidade Europeia de Florença. Foi presidente da Associação Internacional de Inteligência Artificial e Direito. Consultor do Parlamento Europeu em questões de regulação de redes, inteligência artificial e direito digital.