

CONTRIBUIÇÃO DO INTERNETLAB AO EXPEDIENTE T-8.764.298 ACCIÓN DE TUTELA INSTAURADA POR ESPERANZA GÓMEZ SILVA CONTRA FACEBOOK COLOMBIA S.A.S, INSTAGRAM COLOMBIA Y META PLATFORMS, INC, MINISTERIO DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES (MINTIC) Y SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO.

Francisco Brito Cruz, diretor executivo do InternetLab

Fernanda Martins, diretora do InternetLab

Clarice Tavares, coordenadora da área de Desigualdades e Identidades do InternetLab

Iná Jost, coordenadora da área de Liberdade de Expressão do InternetLab

1. ¿Cómo funciona la inteligencia artificial en la moderación de contenidos en plataformas de redes sociales?

Antes de abordar o uso e o funcionamento da inteligência artificial na moderação de conteúdo, é necessário dar um passo atrás e fazer uma breve apreciação sobre o que é moderação de conteúdo em si¹. Consideramos que **essa é uma atividade de elaboração e aplicação de regras**. Com a finalidade de construir diferentes ambientes digitais propícios a interações sociais e geração de conteúdos por seus usuários, cada plataforma decide que tipos de discurso (ou seja, de conteúdo em texto ou multimídia) serão permitidos, incentivados, desincentivados ou proibidos. Para isso, as redes sociais elaboram “*guidelines*”, “*standards*” ou outros conjuntos de regras a serem anexados em seus termos de serviço disponíveis publicamente. Internamente, por sua vez, detalham procedimentos de interpretação de tais regras privadas, organizam precedentes e constroem sistemas de aplicação para dar conta de todo o conteúdo que é gerado por seus usuários.

Todas essas atividades, portanto, fazem parte do referido binômio de funções: elaborar e aplicar as normas que regem e gerenciam o comportamento de usuários de determinado espaço. Administram, portanto, a expressão dessas pessoas, que no caso das redes sociais mais populares produzem quantidades inimagináveis de postagens diariamente².

¹ "Moderação de conteúdo consiste em processo por meio do qual plataformas de internet agem sobre contas ou conteúdos que violem seus termos de uso, impactando sua disponibilidade, visibilidade e/ou credibilidade. A moderação pode envolver diferentes medidas, tais como remoção, suspensão temporária, redução artificial de alcance ou proeminência, superposição de tela de aviso, adição de informação complementar, dentre outras". Thiago Dias Oliva, Victor Pavarin Tavares e Mariana G Valente, “Uma solução única para toda a internet? Riscos do debate regulatório brasileiro para a operação de plataformas de conhecimento”, Diagnósticos e Recomendações (São Paulo: InternetLab, 2020).

² Dados da edição mais recente da pesquisa *Data Never Sleeps* mostram, por exemplo, que a cada minuto, 500 horas de vídeo são postadas no Youtube, e mais de 347 mil tweets são postados. Disponível em: <https://www.domo.com/data-never-sleeps#>



Assim, a apreciação a ser feita sobre o uso de automação na moderação de conteúdo diz respeito ao enorme **desafio logístico que a quantidade de conteúdo que circula nas redes sociais impõe a seus administradores**. Centenas de milhões de “peças” de expressão são diariamente veiculadas no mundo inteiro, em línguas e conjunturas sociais, políticas, culturais e jurídicas diferentes. Entendemos que é impossível prescindir de sistemas automatizados para dar conta de detectar, coletar, organizar tais publicações e atividades online, assim como estabelecer mecanismos efetivos de aplicações de regras privadas de conduta sem no mínimo incorrer em enormes custos e investimentos. A ausência destas ferramentas poderia, inclusive, tornar logisticamente proibitiva a atividade de moderação de conteúdos de grandes plataformas digitais.

Um desafio adicional é a velocidade com a qual se dispersam os conteúdos e se alteram os contextos - e, portanto, os significados do que é potencialmente hostil - configurando, assim, outra dificuldade da moderação. Equívocos nessas tarefas podem ser graves, gerando a supressão de discursos que podem provir de diferentes grupos sociais. E quando são oriundos de pessoas pertencentes a populações historicamente minorizadas, podemos ter o efeito de fortalecimento da invisibilização e silenciamento social que atinge essas populações também fora das mídias sociais.

O campo da inteligência artificial³ soma-se nesse esforço na medida que torna possível “ensinar” computadores a reconhecerem padrões e sinais típicos em conteúdos postados por usuários, entrando em cena quando falamos, por exemplo, das tarefas de detecção, priorização e análise de conteúdos que violam o conjunto de regras das plataformas. Não defendemos com isso que revisões humanas devem ser descartadas, mas sim que seria impossível contar apenas com indivíduos para moderar as atividades dos usuários das plataformas, numa tarefa artesanal. Além de inviabilizar seu serviço, poderíamos gerar um cenário em que as plataformas incorram em mais falhas, dado que não seria possível moderar ativamente todo o conteúdo postado nas redes sociais. Desta forma, entendemos que entre as boas práticas do setor está a composição equilibrada e complementar entre sistemas automatizados (e “inteligentes”) e a participação de colaboradores humanos que supervisionam, revisam e direcionam o trabalho de máquinas.

³ De acordo com Bianca Kremer e Ramon Costa, "a inteligência artificial como conhecemos hoje é baseada majoritariamente em machine learning (aprendizado de máquinas). Isso significa que essas aplicações baseadas em IA são desenvolvidas de modo que as máquinas possam aprender a traçar modelos e resultados sobre problemas que desejamos resolver. A capacidade dessas tecnologias é exponencial, sendo possível que entreguem muitos resultados de forma ágil e em grande volume." Costa, R. S., Kremer, B. Inteligência artificial e discriminação: desafios e perspectivas para a proteção de grupos vulneráveis frente às tecnologias de reconhecimento facial. Revista Brasileira de Direitos Fundamentais & Justiça, 16 (1). Disponível em: <https://dfj.emnuvens.com.br/dfj/article/view/1316/1065>



A inteligência artificial entra, portanto, como artefato fundamental da moderação em escala. De maneira automática, identifica expressões potencialmente infringentes, e as compara com os parâmetros da comunidade, para calcular, contabilizando os fluxos de seu próprio aprendizado, se o conteúdo é violador, e decidindo o que deve ser retirado ou permanecer no ar.

Em resumo, as políticas e termos de uso das plataformas são postas em prática por meio de sistemas que podem conter elementos de inteligência artificial, que, entre outras atividades, possibilita a manutenção de posts legítimos e a derrubada de posts danosos e que vão de encontro às regras. Finalmente, é fundamental encerrar esse capítulo com dois pressupostos de nossas atividades de pesquisa:

- (i) toda inteligência artificial desenvolvida para esta função está apta a cometer erros em suas tarefas de moderação de conteúdo, atuando frequentemente com esquemas que indicam a probabilidade deste ou daquele conteúdo ser infringente;
- (ii) há formas de mitigar erros ou vieses cometidos por inteligência artificial, mas não de eliminá-los completamente, pois estes são decorrentes de realidades mutáveis, complexas e em constante movimento, inerentes da dinâmica da comunicação humana em sociedades altamente conectadas (ponto aprofundado na pergunta 2).

2. ¿Existe evidencia sobre posibles sesgos en las herramientas de inteligencia artificial empleadas para adelantar actividades de moderación de contenidos en redes sociales? Si conoce dicha evidencia, por favor compártala con esta Sala, especialmente sobre sesgos contra las mujeres que desarrollan actividades sexuales pagadas offline o fuera de la plataforma.

Como explicamos acima, toda ferramenta de inteligência artificial poderá vir a cometer erros e apresentar vieses indesejados.⁴⁵ Temos observado o desenvolvimento de pesquisas que, além de demonstrar as possíveis reproduções de desigualdades estruturantes, apontam também para a potencialização de violências que alguns mecanismos de inteligência artificial podem oferecer. Nesse sentido, é importante enfatizar que as tecnologias, o que inclui a inteligência artificial, não são neutras, na medida em que são produzidas e treinadas por pessoas que carregam em si mesmas suas experiências e perspectivas, e também não funcionam de modo estático. Um dos problemas está relacionado às bases de dados para os treinamentos de *machine*

⁴ Para mais informações sobre viés algorítmico, ver em: NOBLE, Safiya Umola. Algorithms of oppression: how search engines reinforce racism. New York: New York University Press, 2018.

⁵ SILVA, Tarcízio. Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código. In: Anais do IV Simpósio Internacional LAVITS – Assimetrias e (In)visibilidades: Vigilância, Gênero e Raça. Salvador, Bahia, Brasil, 2019.



learning. Muitas vezes, as bases de dados utilizadas para treinar o algoritmo para execução de determinadas tarefas possuem limitações - podem ser incompletas ou inviesadas -, e também podem apresentar problemas de atualização - os dados que alimentam uma IA não serão válidos para sempre. Ainda, a inteligência artificial aprende padrões a partir do comportamento de usuários, o que pode gerar a reprodutibilidade de vieses indesejados não só por parte do banco de dados inicial. Por mais sofisticadas e testadas que sejam, com bases de dados complexas e diversas, as tecnologias de inteligência artificial também podem cometer equívocos ao encontrar novas realidades, contextos e problemas que não estavam bem retratados em suas bases de dados de treinamento.

Como seria de imaginar, se toda IA está sujeita a erros e é impossível eliminar totalmente seu risco, temos evidências da detecção de vieses em sistemas de moderação de conteúdo que usam inteligência artificial em temas ligados à sexualidade. Tais evidências demonstram dificuldades próprias enfrentadas pelas plataformas que possuem regras privadas aplicáveis a estes conteúdos, constituindo casos relevantes para refletirmos sobre as estratégias empreendidas pelas empresas de redes sociais para mitigar problemas na sua moderação de conteúdo.

1. Pesquisa conduzida pelo InternetLab testou a ferramenta Perspective, uma tecnologia desenvolvida pela Jigsaw (do grupo Alphabet) que mede o nível de toxicidade de conteúdos em texto, comparando como a tecnologia classificava escritos postados por drag queens e por nacionalistas brancos. A pesquisa aponta que este tipo de tecnologia pode não ser capaz de distinguir nuances de contextos, como por exemplo, a diferença entre discursos de ódio contra pessoas LGBTQIA+ e conteúdo publicado por pessoas LGBTQIA+, que frequentemente ressignificam termos compreendidos como ofensivos, positivando-os para comunicação entre pares. Os resultados indicaram que os perfis de drag queens foram considerados potencialmente mais tóxicos do que os perfis de supremacistas brancos. Termos comumente usados por pessoas LGBTQIA+, como “sissy”, “gay”, “lesbian” e “queer” eram considerados potencialmente tóxicos pela ferramenta de IA, independentemente de contexto.⁶ O caso da pesquisa sobre *drag queens* demonstra distorções e vieses da inteligência artificial, que não foi capaz de identificar as matizes específicas da comunicação de pessoas *queer*, em que o uso de uma linguagem

⁶ Dias Oliva, T., Antonialli, D.M. & Gomes, A. Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture* 25, 700-732 (2021). <https://doi.org/10.1007/s12119-020-09790-w>; and InternetLab. “Drag queens e Inteligência Artificial: computadores devem decidir o que é ‘tóxico’ na internet?”. 28 de junho de 2019. Disponível em: <https://internetlab.org.br/pt/noticias/drag-queens-e-inteligencia-artificial-computadores-devem-decidir-o-que-e-toxico-na-internet/>.



“pseudo-ofensiva” é uma forma comum de comunicação entre membros da própria comunidade.⁷ Ainda, este exemplo sugere que as bases de dados usadas para treinar a inteligência artificial eram inconsistentes ou parciais, já que o algoritmo interpretava como tóxicos termos que, em si, não são necessariamente violentos ou ofensivos.

2. Outro exemplo que ilustra as limitações das tecnologias de inteligência artificial refere-se aos casos de moderação imprecisa e/ou equivocada em publicações com nudez. Algumas plataformas proíbem, de forma legítima, a disseminação de imagens contendo partes de um corpo desnudo, como, por exemplo, a Meta. Isso ocorre não só por decisões de negócio sobre qual o tipo de conteúdo erótico é ou não aceitável, mas também porque uma escolha como essa pode reduzir riscos de circulação de conteúdos de exploração sexual de crianças e adolescentes pela dificuldade de aferição de idade precisa em imagens e vídeos.⁸ Em geral, quando há previsão do banimento desse tipo de conteúdo, são elencadas como exceções a essas políticas imagens com nudez que façam referência a (i) momentos de parto e pós-parto, (ii) procedimentos cirúrgicos e médicos de confirmação de gênero, e (iii) autoexames de câncer ou publicações sobre prevenção e avaliação de doenças em genitálias e partes íntimas.⁹ No entanto, existem casos recorrentes de remoção de conteúdo, especialmente de pessoas trans e não-binárias, com publicações sobre procedimentos de afirmação de gênero, com base nas políticas de nudez e de proposta de cunho sexual. Em resposta às remoções, pessoas trans e não-binárias engajaram-se no movimento #DeserveToBeHere (em tradução livre, #MerecemosEstarAqui), em que contestavam banimento de fotos com nudez sobre transição de gênero.¹⁰

⁷ Murray, Stephen O. “The Art of Gay Insulting.” *Anthropological Linguistics*, vol. 21, no. 5, 1979, pp. 211–23. JSTOR, <http://www.jstor.org/stable/30027635>.

⁸ Em resposta ao Comitê de Supervisão do Facebook, a Meta afirmou, por exemplo, que na elaboração dos princípios gerais de suas políticas sobre nudez, a empresa considerou: “(1) a natureza privada ou sensível das imagens; (2) se foi dado consentimento para a obtenção e compartilhamento de imagens de nudez; (3) o risco de exploração sexual; e (4) se a divulgação das imagens pode levar a assédio fora da plataforma, particularmente em países onde elas podem ser culturalmente ofensivas”. Disponível em: <https://oversightboard.com/decision/BUN-IH313ZHJ/>

⁹ Política de Nudez adulta e atividades sexuais da Meta. Disponível em:

<https://transparency.fb.com/en-gb/policies/community-standards/adult-nudity-sexual-activity/>.

¹⁰ Disponível em:

<https://www.thepinknews.com/2021/04/22/instagram-trans-bodies-censorhip-we-deserve-to-be-here/>



INTERNETLAB

O tema da identidade de gênero e nudez foi objeto de questionamento no âmbito do Comitê de Supervisão do Facebook¹¹, ao qual o InternetLab enviou um comentário. No caso, um casal de pessoas trans questionava a remoção de fotos em que estavam com seios expostos e mamilos cobertos com uma fita, em uma campanha de arrecadação de dinheiro para cirurgia de afirmação de gênero.¹² Em sua decisão, o próprio Comitê revogou a decisão da Meta de exclusão das fotos e reforçou a necessidade de revisão e de melhora nas orientações sobre moderação de conteúdo, como uma forma de reduzir erros de remoções.¹³ Assim, um aprimoramento da análise de contexto se faz essencial para moderação de conteúdo, especialmente em casos envolvendo grupos historicamente marginalizados, que são afetados de forma desproporcional.

Contudo, como dito acima, é importante ressaltar que ainda que existam casos de erros e discriminações, bem como limitações, o uso de inteligência artificial para a moderação de conteúdo em plataformas de redes sociais é inevitável. O volume de postagens, diversidade de línguas e contextos torna impossível que a aplicação das regras das plataformas seja feita, exclusivamente, por humanos. Dados da pesquisa DataNeverSleep de 2020, conduzido pela Domo,¹⁴ apontam que, a cada minuto, são postados 347 mil novos stories no Instagram e são postadas 147 mil fotos no Facebook.¹⁵ Assim, em um contexto em que o uso de ferramentas de inteligência artificial é imprescindível, dado o desafio logístico de moderar milhares de conteúdos publicados por minuto, é preciso que essas ferramentas estejam em um constante aprimoramento. Devem contar com bases de dados mais completas e complexas, uma melhor atenção aos contextos e às

¹¹ O Comitê de Supervisão do Facebook é órgão independente responsável por revisar decisões de moderação de conteúdo do Facebook e do Instagram e dar recomendações quanto às políticas de moderação de conteúdo na plataforma. Em seu website, o Comitê afirma que o seu objetivo é “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.” Disponível em: <https://www.oversightboard.com/>.

¹² Para mais informações, ver contribuição do InternetLab ao o caso 2022-009-IG-UA & 2022-010-IG-UA do Comitê de Supervisão do Facebook. MARTINS, Fernanda; TAVARES, Clarice; FERREIRA, Lux. “Comentário do InternetLab sobre o caso 2022-009-IG-UA & 2022-010-IG-UA do Comitê de Supervisão do Facebook”. Disponível em:

<https://internetlab.org.br/pt/noticias/internetlab-envia-contribuicao-para-caso-de-identidade-de-genero-e-nudez-do-comite-de-supervisao-do-facebook/>.

¹³ Decisão do Comitê de Supervisão do Facebook está disponível em: <https://www.oversightboard.com/news/1214820616135890-oversight-board-overturms-meta-s-original-decisions-in-the-gender-identity-and-nudity-cases/>

¹⁴ Domo é uma empresa de software em nuvem que publica, anualmente, o DataNeverSleep, estudo em que apresenta o número de conteúdos postado em cada plataforma por minuto. Disponível em:

<https://www.domo.com/>.

¹⁵ TechTudo. “O que acontece a cada minuto na Internet? Estudo traz dados surpreendentes”. 14 de agosto de 2020. Disponível em:

<https://www.techtudo.com.br/noticias/2020/08/o-que-acontece-a-cada-minuto-na-internet-estudo-traz-dados-surpreendentes.ghtml>. Os dados completos da pesquisa da Domo estão disponíveis em:

<https://www.domo.com/data-never-sleeps#>.



especificidades locais e de grupos historicamente marginalizados. Além disso, como será desenvolvido, a inteligência artificial deve ser combinada a outros recursos para a moderação de conteúdo, já que a máquina sozinha - em seu atual estado de desenvolvimento - não é capaz de detectar e compreender todas as particularidades da comunicação humana.

3. ¿Qué estrategias pueden implementarse para superar los posibles sesgos de las herramientas de inteligencia artificial empleadas en la moderación de contenidos en redes sociales? Si conoce evidencia al respecto, por favor compártala con esta Corporación judicial.

Assumidas as premissas de que o exercício da moderação de conteúdo pressupõe o uso de inteligência artificial, bem como que essas ferramentas geram vieses que impactam negativamente a gestão da expressão online, passamos a discorrer sobre mecanismos que podem auxiliar a mitigar essas consequências.

A primeira vertente de ações de atenuação desses riscos diz respeito a **estratégias de transparência sobre os mecanismos de moderação, combinada com métodos de prestação de contas e regras sobre devido processo na moderação de conteúdo**. Para começar, é fundamental que as normas e termos de uso que regem as plataformas sejam escritas em linguagem clara e acessível, traduzidas aos idiomas de todas as regiões onde são navegadas, e dispostas de maneira que qualquer pessoa possa encontrá-los. Em seguida, que existam métodos de prestação de contas, isto é, que as pessoas que acessam as redes sociais possam dar opiniões e *feedbacks* sobre o exercício da moderação. Essa ferramenta tem o potencial de gerar um ciclo virtuoso de melhora, já que a inteligência artificial bebe de sua própria fonte, aprimorando-se com o uso. Dessa forma, a aplicação dos termos de uso e o rendimento do sistema estariam, de fato, sendo constantemente avaliados. Por fim, que as plataformas possuam instâncias recursais, de maneira a possibilitar que as pessoas possam, por meio de procedimentos claros e formais, conhecer a razão das decisões que influenciam sua expressão nas redes, e contestá-las, caso julguem justo e necessário.

O segundo âmbito de ações fundamentais para atenuar deslizos da automação é a incorporação de etapas de supervisão realizadas por pessoas qualificadas que revisem e direcionem o trabalho das máquinas. Apesar de concordarmos que a automação por meio do uso de inteligência artificial é indispensável, consideramos que esses processos precisam ser complementados com recursos humanos que os aprimorem e agreguem nuances às suas análises, freios e contrapesos para abrandar eventuais vieses. Tal revisão adiciona matizes às decisões maquinadas, e somam especificidades e contextos que podem ser peças-chave para uma gestão saudável do discurso que circula



nas plataformas. Isso é especialmente importante quando a possibilidade de cometer equívocos afeta usuários desproporcionalmente, como em casos de populações historicamente minorizadas cujo discurso já sofre limitações fora do ambiente online. O [InternetLab aborda essa questão em pesquisa que discorre sobre sistemas de moderação de conteúdo em camadas¹⁶](#).

Por fim, pontua-se que sistemas como esses precisam de constante investimento e reavaliação. Diante da enorme quantidade de conteúdo, de sua velocidade de circulação e das frequentes mudanças de contextos que envolvem a expressão, é necessário que as plataformas dediquem um olhar cuidadoso e continuado para essa questão. Isso deve envolver, por exemplo, diálogo com especialistas locais, equipes interdisciplinares com pontos focais regionais atentos e especializados, capazes de abordar desafios atrelados a idiomas e conjunturas específicos, como um processo eleitoral, a promulgação de determinada legislação, ou o início de um conflito. Além disso, é essencial que existam métricas específicas e transparentes de análise, e que todos esses mecanismos combinados sejam capazes de fomentar as melhorias nas ferramentas tecnológicas, e, por consequência, na gestão da moderação de conteúdo.

4. Por último, según su conocimiento del funcionamiento de las plataformas de red social ¿cómo sus operadores previenen convertirse en intermediarios o gatekeepers de otras plataformas o sitios web con contenido pornográfico? Por ejemplo, frente al uso de enlaces a plataformas con contenido para adultos explícito como sitios web de pornografía o redes sociales propias del negocio webcam (ej. OnlyFans)

O InternetLab não possui pesquisas específicas sobre o papel de intermediários ou “gatekeepers” exercido por plataformas de redes sociais em relação ao mercado de produção de conteúdo adulto ou o ecossistema de distribuição de pornografia lícita online. Contudo, sobre este tema, conseguimos formular ao menos dois pontos que precisam ser considerados para uma reflexão acerca do papel das plataformas enquanto intermediárias de conteúdo pornográfico.

O primeiro ponto é que se tratando de conteúdo adulto, há camadas extras de sensibilidade em razão do risco de atividades extremamente nocivas aproveitarem brechas e permissões, como na produção e disseminação de conteúdo de exploração

¹⁶ A moderação em camadas traz etapas adicionais de análise qualificada para determinados tipos de perfis ou conteúdos. Essa ferramenta pode ser uma estratégia empregada pelas empresas para mitigar riscos aos direitos humanos, uma vez que dão prioridade de análise a alguns tipos de usuários ou posts que devem ser cuidadosamente revisados visando proteger certos tipos de discurso, como por exemplo, expressões de populações historicamente marginalizadas, ativistas políticas, jornalistas, entre outras.

[O estudo também está disponível em versão em inglês.](#)



sexual, o que pode incluir não só pessoas adultas, mas também crianças e adolescentes, ou na divulgação não consentida de imagens íntimas.

Um segundo ponto a ser ressaltado é que **as plataformas de redes sociais são, por excelência, intermediárias de conteúdos de terceiros**. Ou seja, não são autoras dos conteúdos que são disponibilizados nas redes sociais, senão que hospedam conteúdos de outros indivíduos. Enquanto intermediárias, as plataformas têm a responsabilidade de determinar regras, por meio de suas políticas internas, para estabelecer quais são as atividades legítimas ou ilegítimas, com base no modelo de serviço online que pretendem oferecer. A capacidade de desenvolver diferentes regras, com diferentes formas de moderação de conteúdo, é chave para que os espaços criados online possuam vocações próprias, criando uma internet plural e diversa, em que cada uma das plataformas possa ter uma identidade própria, com usuários e interesses específicos.¹⁷

Essa diversidade de modelos de serviço pode ser exemplificada com as diferentes formas como as plataformas lidam com políticas sobre nudez. Enquanto algumas plataformas como o X, antigo Twitter, permitem nudez e em que o usuário pode esperar ter acesso a este tipo de conteúdo, outras, como Facebook e Instagram, vedam este tipo de conteúdo, como característica dos ambientes online que viabilizam o seu modelo de negócio. Ainda, existem casos como o Tumblr, rede social utilizada especialmente para divulgação de conteúdos artísticos, que alterou suas políticas de nudez. Em 2018, o Tumblr deixou de permitir conteúdo sexual explícito ou contendo nudez, para evitar o banimento do seu aplicativo da App Store e para melhor se adequar as políticas dos processadores de pagamento e do servidores.¹⁸ Mais recentemente, a plataforma voltou a permitir apenas conteúdos com nudez, sem atividades sexuais.¹⁹ As alterações das regras do Tumblr evidenciam que tais plataformas podem mudar suas políticas no tempo e que há uma miríade de relações a serem consideradas entre diferentes tipos de intermediários, não só entre plataformas de redes sociais e sites web.

¹⁷ Artur Pericles Lima Monteiro, Francisco Brito Cruz, Juliana Fonteles da Silveira e Mariana G. Valente, “Armadilhas e caminhos na regulação da moderação de conteúdo”, Diagnósticos & Recomendações (São Paulo: InternetLab, 2021). Disponível em:

https://internetlab.org.br/wp-content/uploads/2021/09/internetlab_armadilhas-caminho-moderacao.pdf.

¹⁸ The Verge. “Tumblr will now allow nudity but not explicit sex”. Available at:

<https://www.theverge.com/2022/11/1/23435516/tumblr-porn-ban-modified-nudity-allowed-sexually-explicit-depictions-banned>.

¹⁹ Em sua atual política, a plataforma afirma que “nudez e outros tipos de conteúdo adulto em geral são bem-vindos. Não estamos aqui para julgar sua arte, apenas pedimos que você adicione um aviso de conteúdo para que as pessoas possam optar por filtrá-lo do painel, se preferirem.” Disponível em:

<https://www.tumblr.com/policy/br/community>.



A multiplicidade de regras e práticas de moderação no que tange a conteúdo sexual e de nudez dinamiza a produção e disseminação de conteúdo em diferentes plataformas, permitindo ao usuário uma diversidade de formas de relacionamento e de usos dos serviços de redes sociais. Em uma plataforma que permita conteúdo com nudez, o usuário tem a expectativa de encontrar uma determinada linguagem, mais voltada ao público adulto; enquanto naquelas em que este tipo de conteúdo é proibido, espera-se que o tipo de interação e linguagem sejam apropriadas a todas as idades.²⁰ Assim, **cada plataforma tem espaço e liberdade para desenhar suas políticas de formas diferentes, como uma forma de reforçar a liberdade de expressão, na medida em que essa multiplicidade possibilita a criação de espaços que comportam expressões diferentes, fomentando a pluralidade dos discursos.**

²⁰ Artur Pericles Lima Monteiro, Francisco Brito Cruz, Juliana Fonteles da Silveira e Mariana G. Valente, “Armadilhas e caminhos na regulação da moderação de conteúdo”, Diagnósticos & Recomendações (São Paulo: InternetLab, 2021). Disponível em: https://internetlab.org.br/wp-content/uploads/2021/09/internetlab_armadilhas-caminho-moderacao.pdf.

