IV CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL (IV CIDIA)

DIREITO CIBERNÉTICO, LIBERDADE DE EXPRESSÃO E PROTEÇÃO DE DADOS II

D598
Direito cibernético, liberdade de expressão e proteção de dados II [Recurso eletrônico on-line] organização IV Congresso Internacional de Direito e Inteligência Artificial (IV CIDIA): Skema Business School – Belo Horizonte;

Coordenadores: Aghisan Xavier Ferreira Pinto, Marina de Castro Firmo e Luiza Santos Cury Soares – Belo Horizonte: Skema Business School, 2023.

Inclui bibliografia
ISBN: 978-65-5648-777-9
Modo de acesso: www.conpedi.org.br em publicações
Tema: Os direitos dos novos negócios e a sustentabilidade.
1. Direito. 2. Inteligência artificial. 3. Tecnologia. I. IV Congresso Internacional de Direito e Inteligência Artificial (1:2023 : Belo Horizonte, MG).

CDU: 34

IV CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL (IV CIDIA)

DIREITO CIBERNÉTICO, LIBERDADE DE EXPRESSÃO E PROTEÇÃO DE DADOS II

Apresentação

O IV Congresso Internacional de Direito e Inteligência Artificial - CIDIA da SKEMA Business School Brasil, realizado nos dias 01 e 02 de junho de 2023 em formato híbrido, consolida-se como o maior evento científico de Direito e Tecnologia do Brasil. Estabeleceram-se recordes impressionantes, com duzentas e sessenta pesquisas elaboradas por trezentos e trinta e sete pesquisadores. Dezenove Estados brasileiros, além do Distrito Federal, estiveram representados, incluindo Amazonas, Bahia, Ceará, Distrito Federal, Espírito Santo, Goiás, Maranhão, Minas Gerais, Pará, Pernambuco, Paraná, Rio de Janeiro, Rio Grande do Norte, Rondônia, Roraima, Rio Grande do Sul, Santa Catarina, Sergipe, São Paulo e Tocantins.

A condução dos trinta e três grupos de trabalho do evento, que geraram uma coletânea de vinte e cinco livros apresentados à comunidade científica nacional e internacional, contou com a valiosa colaboração de sessenta e três professoras e professores universitários de todo o país. Esses livros são compostos pelos trabalhos que passaram pelo rigoroso processo de double blind peer review (avaliação cega por pares) dentro da plataforma CONPEDI. A coletânea contém o que há de mais recente e relevante em termos de discussão acadêmica sobre a relação entre inteligência artificial, tecnologia e temas como acesso à justiça, Direitos Humanos, proteção de dados, relações de trabalho, Administração Pública, meio ambiente, sustentabilidade, democracia e responsabilidade civil, entre outros temas relevantes.

Um sucesso desse porte não seria possível sem o apoio institucional de entidades como o CONPEDI - Conselho Nacional de Pesquisa e Pós-graduação em Direito; o Programa RECAJ-UFMG - Ensino, Pesquisa e Extensão em Acesso à Justiça e Solução de Conflitos da Faculdade de Direito da Universidade Federal de Minas Gerais; o Instituto Brasileiro de Estudos de Responsabilidade Civil - IBERC; a Comissão de Inteligência Artificial no Direito da Ordem dos Advogados do Brasil - Seção Minas Gerais; a Faculdade de Direito de Franca - Grupo de Pesquisa Políticas Públicas e Internet; a Universidade Federal Rural do Semi-Árido - UFERSA - Programa de Pós-graduação em Direito - Laboratório de Métodos Quantitativos em Direito; o Centro Universitário Santa Rita - UNIFASAR; e o Programa de Pós-Graduação em Prestação Jurisdicional e Direitos Humanos (PPGPJDH) - Universidade Federal do Tocantins (UFT) em parceria com a Escola Superior da Magistratura Tocantinense (ESMAT).

Painéis temáticos do congresso contaram com a presença de renomados especialistas do Direito nacional e internacional. A abertura foi realizada pelo Professor Dierle Nunes, que discorreu sobre o tema "Virada tecnológica no Direito: alguns impactos da inteligência artificial na compreensão e mudança no sistema jurídico". Os Professores Caio Lara e José Faleiros Júnior conduziram o debate. No encerramento do primeiro dia, o painel "Direito e tecnologias da sustentabilidade e da prevenção de desastres" teve como expositor o Deputado Federal Pedro Doshikazu Pianchão Aihara e como debatedora a Professora Maraluce Maria Custódio. Para encerrar o evento, o painel "Perspectivas jurídicas da Inteligência Artificial" contou com a participação dos Professores Mafalda Miranda Barbosa (Responsabilidade pela IA: modelos de solução) e José Luiz de Moura Faleiros Júnior ("Accountability" e sistemas de inteligência artificial).

Assim, a coletânea que agora é tornada pública possui um inegável valor científico. Seu objetivo é contribuir para a ciência jurídica e promover o aprofundamento da relação entre graduação e pós-graduação, seguindo as diretrizes oficiais da CAPES. Além disso, busca-se formar novos pesquisadores na área interdisciplinar entre o Direito e os diversos campos da tecnologia, especialmente o da ciência da informação, considerando a participação expressiva de estudantes de graduação nas atividades, com papel protagonista.

A SKEMA Business School é uma entidade francesa sem fins lucrativos, com uma estrutura multicampi em cinco países de diferentes continentes (França, EUA, China, Brasil e África do Sul) e três importantes acreditações internacionais (AMBA, EQUIS e AACSB), que demonstram sua dedicação à pesquisa de excelência no campo da economia do conhecimento. A SKEMA acredita, mais do que nunca, que um mundo digital requer uma abordagem transdisciplinar.

Expressamos nossos agradecimentos a todas as pesquisadoras e pesquisadores por sua inestimável contribuição e desejamos a todos uma leitura excelente e proveitosa!

Belo Horizonte-MG, 14 de julho de 2023.

Profª. Drª. Geneviève Daniele Lucienne Dutrait Poulingue

Reitora – SKEMA Business School - Campus Belo Horizonte

Prof. Dr. Caio Augusto Souza Lara

Coordenador de Pesquisa – SKEMA Law School for Business

LIBERDADE DE EXPRESSÃO NAS REDES SOCIAIS E A IMPLEMENTAÇÃO DA MODERAÇÃO DE CONTEÚDO POR MEIO DE INTELIGÊNCIA ARTIFICIAL

FREEDOM OF SPEECH ON SOCIAL MEDIA AND THE IMPLEMENTATION OF CONTENT MODERATION THROUGH ARTIFICIAL INTELLIGENCE

Lucas Matheus Dutra Bandeira [1]
João Carlos Bezerra de Araújo Costa e Silva [2]
Rosana dos Santos Martins [3]

Resumo

O presente artigo buscou apresentar a liberdade de expressão no cenário de aplicação das novas tecnologias moderadoras de conteúdo, objetivando a análise do uso de IAs em redes sociais e seus respectivos impactos no direito fundamental à liberdade de expressão, ante a postagens que infrinjam a tolerância social e padrões de postagens de respectivas comunidades online, assim como a limitações e desafios enfrentados em seu processo de implementação.

Palavras-chave: Inteligência artificial, Moderação de conteúdos, Liberdade de expressão

Abstract/Resumen/Résumé

This article sought to present the freedom of expression in the scenario of application of new technologies for content moderation, aiming to analyze the use of AIs in social networks and their respective impacts on the fundamental right to freedom of expression, in the face of posts that violate social tolerance and posting standards of respective online communities, as well as limitations and challenges faced in their implementation process.

Keywords/Palabras-claves/Mots-clés: Artificial intelligence, Content moderation, Freedom of expression

[1] Estudante Graduando de Direito - UNIFOR. Membro do Grupo de Estudos em Tecnologia, Informação e Sociedade - GETIS

[2] Estudante Graduando em Direito - UNIFOR. Membro do Grupo de Estudos em Tecnologia, Informação e Sociedade - GETIS

[3] Advogada. Mestre em Desenvolvimento Social - Unimontes. Membro do Grupo de Estudos em Tecnologia, Informação e Sociedade – GETIS

INTRODUÇÃO

O advento das novas tecnologias e a ascensão das mídias digitais, trouxe consigo a capacidade amplificativa dos discursos nos meios sociais, já que por meio da internet, houve o assentamento da praticidade integrativa de pessoas em diversos locais ao redor do globo (HARTMANN; SILVA, 2022). Sob esta perspectiva, a liberdade de expressão no ambiente digital tornou-se um tema recorrente ante aos questionamentos e aos receios das aplicações dos usos de Inteligências Artificiais (IA) como novos modelos moderadores das liberdades de falas nas redes, já que tal utilização tem como objetivo a praticidade e aplicabilidade de responsabilização pelos discursos no ambiente digital.

O presente trabalho tem como objetivo analisar os usos de moderadores IAs nas redes sociais, buscando correlacionar o perfil contemporâneo do direito fundamental à liberdade de expressão e pensamento com a análise dos respectivos e possíveis impactos que a moderação artificial eventualmente proporciona no modo de se expressar humano no ambiente digital.

MATERIAIS E MÉTODOS

A metodologia empregada no presente trabalho foi a dedutiva, esta que consiste em utilizar um modelo lógico-argumentativo empírico onde, partindo de pressupostos gerais e utilizando-se da lógica, busca-se chegar às conclusões (HENRIQUES e MEDEIROS, 2015; MICHEL, 2017). Para a forma de abordagem do problema, foi utilizada a qualitativa, tendo em conta que representação quantitativa não se confirmou no foco e sentido deste trabalho, visto que o modelo qualitativo baseia-se essencialmente na coleta e entendimento dos dados utilizados como referenciais teóricos para a análise do tema a ser estudado (HENRIQUES e MEDEIROS, 2015; MICHEL, 2017). Segundo a coleta de dados, a pesquisa foi do tipo bibliográfica e sistemática, onde foram utilizadas publicações referentes ao tema nomeadamente artigos, livros, monografias, dissertações e teses de doutorado, com o fito de melhor entender as nuances dos assuntos estudados que o resumo expandido se propôs a investigar e relacionar.

RESULTADOS

A moderação de conteúdo não é uma ferramenta exclusiva do novo século e limitada à aplicação nas mídias sociais, muito pelo contrário, sendo utilizada desde as primeiras formas de comunicação e muito presente na imprensa tradicional. Basta trazer para análise a disposição das notícias em jornais e revistas, onde estas eram expostas ao crivo de departamentos de curadoria especializados em decidir as matérias que seriam ou não publicadas e qual o destaque

que estas teriam de acordo com a relevância que julgavam possuir, ou que desejavam que possuíssem.

Entretanto, na realidade em que vivemos, as notícias de maior alcance são aquelas veiculadas nas mídias sociais, sendo estas apenas o veículo pelo meio do qual os usuários as publicam, não funcionando como um produtor de conteúdo, como era o caso das mídias tradicionais. Tal realidade, impossibilita que estas plataformas realizem um filtro de moderação prévio à publicação.

Ainda nesse cenário, frente à ausência deste controle publicitário, as plataformas sociais realizam, com base nas suas diretrizes e padrões terminados, o controle de forma posterior às publicações, seja de forma proativa, mediante aplicação de ferramentas dotadas de inteligência artificial e empresas de moderação de conteúdo, ou motivada, que ocorre mediante auxílio da própria comunidade por meio de sinalizações e denúncias.

Nesse esteio, cabe pontuar que, alguns são os fatores, segundo Sander (2020), que impactam na forma como a moderação ocorre nas redes sociais, como é o caso da própria filosofia da empresa, onde, as diversas plataformas existentes no mercado possuem públicos e objetivos distintos, ao passo que algumas prezam por um ambiente de livre expressão e pouca, ou nenhuma, intervenção, outras buscam manter mais acessível e amigável ao público em geral.

Nesse prisma, vale ressaltar que estas plataformas não funcionam em ambientes alheios à realidade e às leis, razão pela qual precisam se adequar às exigências regulatórias dos países em que ofereçam seus serviços, o que impacta na forma como lidam com a moderação dos conteúdos publicados, como exemplo há a implementação do Código de Conduta para Combate ao Discurso de Ódio pela UE em parceria com diversas plataformas.

Como comumente, e com objetivos cômicos, mas não totalmente equivocados, é citado que "a função social da empresa é gerar lucro", assim também é motivada a moderação das empresas responsáveis pelas redes sociais, objetivando manter o sistema agradável aos seus usuários para que estes se mantenham ativos, fator que está diretamente ligado ao ponto seguinte.

Assim, o último dos pontos elencados por Sander, a opinião pública, é capaz de resumir a generalidade dos demais, uma vez que estas plataformas precisam manter seus usuários, atitudes e conteúdos tidos por imorais por uma sociedade, caso infestem as plataformas, motivarão a debandada de membros, assim como a moderação de conteúdos de forma equivocada também, como é o caso do Facebook, que após retirar várias postagens de mulheres amamentando motivou protestos de insatisfação da população.

Além disso, protestos como este levantam o questionamento de quais são os níveis de intervenção destes mecanismos de moderação, como estes ocorrem, quem ou o que é o responsável pelos dados que fundamentaram as decisões tomadas e fomentam as reflexões acerca da intervenção das plataformas nos conteúdos publicados por seus usuários.

Dentro das plataformas de relações sociais, alguns métodos de moderação são mais comuns e possuem sua aplicação empregada com maior recorrência, como por exemplo, a remoção do material, aplicado aos casos de publicação de conteúdos ilícitos, como incitação ao terrorismo, e consiste na eliminação completa do conteúdo postado.

Nessa conjuntura, recorre-se também à indisponibilização, que se trata de uma técnica de indisponibilização do conteúdo temporária ou geograficamente, como é o caso do *geoblocking*, ou à restrição, que ocorre por meio de uma ocultação parcial da publicação e inserção de mensagem informando do teor do conteúdo e se o usuário deseja realmente seguir com a visualização deste.

Sob esse viés, ainda é utilizada a prática do ranqueamento, similar à curadoria das mídias tradicionais, onde o algoritmo irá priorizar o conteúdo mais relevante, replicando-o em detrimento daqueles com menor relevância. O grande desafio da implementação das tecnologias de inteligência artificial no ambiente das redes sociais para a moderação de conteúdo é sua capacidade de analisar e entender fatores sociais, culturais, históricos e políticos.

Por outro lado, ao passo que as ferramentas dotadas de *machine learning* são excelentes analistas de padrões, é inegável que haja uma complexidade latente na análise de vídeos, demandando a averiguação de múltiplos quadros e dados em áudio, no caso dos comuns "memes" a tarefa se torna ainda mais complexa, uma vez que demanda da inteligência artificial a compreensão do contexto subjetivo daquela mídia. Ambas as análises podem ser agravadas quando temos a publicação ao vivo, por meio de *livestreams*, que cobram da IA todas essas capacidades cognitivas em tempo real.

Estas dificuldades são notadas comumente ao verificar postagens de cunho inofensivo, como uma enquete literária, sendo restringidas por serem confundidas com conteúdo eleitoral por possuírem as palavras "votação" e "vencedor", charges incitando a violência passando despercebidas pelo algoritmo ou censura de imagens de estátuas postadas por turistas por conterem "nudez e conteúdo sexual".

Com base nesse quadrante, tais equívocos provocados pela IA ao avaliar os conteúdos veiculados nas redes sociais podem impactar diretamente no senso de verdade dos usuários, atuando passivamente como um limitador da liberdade de expressão dos indivíduos participantes, tendo em conta que ao priorizar determinadas matérias sob um espectro

específico, por, segundo números, ser este o que mais agrada o público, há a limitação indireta do alcance de divergências discursivas, o que resulta em nichos que tomam determinadas opiniões como uma verdade universal para aqueles que utilizam a mídia social.

A função mais conhecida do uso da inteligência artificial na moderação de conteúdo é exatamente a pré-moderadora, onde a IA realiza a primeira análise do conteúdo publicado nas redes sociais, verificando se há algum dos sinalizadores de atenção (conteúdo sexual, informação falsa, violência, etc) e verificando se este pode ser mantido em circulação, se necessita ser restringido ou se demanda uma segunda análise por moderador humano, encaminhando este conteúdo para análise de empresas especializadas nessa atividade. Uma vez analisado o conteúdo pelo moderador, essas informações amplificam a base de dados da IA.

Além desta, as tecnologias de inteligência artificial também auxiliam na tarefa dos moderadores, aplicando filtros e sinalizações para que estes estejam mais cientes do que visualizarão e possam mitigar os danos causados aos profissionais, como os recorrentes casos de síndrome do pânico, crises de ansiedade, depressão e aversão ao contato social, observados após as jornadas de exposição a esses conteúdos.

Outra utilização é o uso da IA para treinamento, onde uma inteligência artificial é programada para gerar conteúdos reprováveis para que possam ser utilizados no aprimoramento do algoritmo de análise. Diversas são as utilizações e os respectivos impactos alcançados pelo uso da IA na tarefa de moderação de conteúdo dentro das plataformas sociais, os avanços são muitos, a humanidade muito se beneficia desta tecnologia em sua rotina ou mesmo passo que sofre com sua interferência na forma como se comunica e goza de seu direito de liberdade de expressão. Ambas as inteligências se propulsionam para o desenvolvimento mútuo.

CONSIDERAÇÕES FINAIS

A presente pesquisa consistiu na identificação das implicações da moderação de conteúdo nas -redes sociais frente à liberdade de expressão, uma vez que aquelas fazem uso da IA como um filtro para postagens que infrinjam os padrões da comunidade.

O controle de conteúdo fundamenta-se nos padrões e diretrizes das redes sociais e pode acontecer por uma análise posterior da plataforma por meio da IA ou de empresas especializadas. Entretanto, a moderação de conteúdo é uma atividade complexa, uma vez que, a partir das análises, a plataforma decide se tal conteúdo pode ou não estar no ambiente digital, muitas vezes de forma errônea, limitando a liberdade de expressão do usuário. Assim, as postagens dos usuários podem ser passíveis de remoção, indisponibilização, restrição ou ranqueamento após as análises.

Umas das formas de moderação de conteúdo se dá pelo uso da IA, em especial as ferramentas dotadas de *machine learning* na fase de pré-moderação. Seu uso garante uma celeridade nas análises, porém há limitações que precisam ser consideradas, sendo um dos grandes desafios a dificuldade da IA em entender aspectos culturais, linguísticos, políticos e históricos fazendo com que postagens sejam suspensas ou restringidas com base em uma análise errônea da máquina.

Portanto, depreende-se, que é inegável que a IA possui um papel de grande valia na moderação de conteúdo, porém, faz-se necessário uma maior atenção ao seu uso para não cairmos em uma censura privada por meio de uma remoção arbitrária de conteúdo.

REFERÊNCIAS

CAMBRIDGE CONSULTANTS. **Use of AI in online content moderation**. 2019. Disponível em: https://www.epra.org/news_items/ofcom-s-report-the-use-of-ai-in-content-moderation. Acesso em 15 nov. 2022

DELGROSSI, Sara; SAID-MOORHOUSE, Lauren. Facebook banned Neptune statue photo for being 'explicitly sexual'. **CNN**, 5 janeiro 2017. Disponível em: https://edition.cnn.com/travel/article/facebook-neptune-statue-photo-ban/index.html. Acesso em: 14 nov. 2022.

EUROPEAN UNION. **Code of conduct on countering illegal hate speech online**, 2016. Disponível em:https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en#theeucodeofconduct. Acesso em: 12 nov. 2022.

HARTMANN, I. A.; SILVA, L. A. da. Inteligência artificial e moderação de conteúdo: o sistema CONTENT ID e a proteção dos direitos autorais na plataforma Youtube. **IUS GENTIUM**, [*S. l.*], v. 10, n. 3, p. 145–165, 2020. DOI: 10.21880/ius gentium.v10i3.503. Disponível em: https://www.revistasuninter.com/iusgentium/index.php/iusgentium/article/view/503. Acesso em: 30 nov. 2022.

HENRIQUES, A.; MEDEIROS, J.B. **Metodologia Científica da Pesquisa Jurídica**, 9ª edição. São Paulo Grupo GEN, 2017. 9788597011760. Disponível em: https://integrada.minhabiblioteca.com.br/#/books/9788597011760/. Acesso em: 30 nov. 2022

SANDER, Barrie. Freedom of Expression in the Age of On-line Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content. **Fordham International Law Journal**, v. 43:4, p. 948-954,2020.Disponível em:Moderation.https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2787&context=ilj. Acesso em: 30 mar. 2023.

SWENEY, Mark. Mums furious as Facebook removes breastfeeding photos. **The Guardian**, 30 de dezembro de 2008. Disponível em: https://www.theguardian.com/media/2008/dec/30/facebook-breastfeeding-ban. Acesso em: 11 nov. 2022.