

Mineração de textos e dados na pesquisa em saúde: reflexões sobre direitos autorais

Text and data mining in health research: reflections on copyright

Minería de textos y datos en la investigación en salud: reflexiones sobre los derechos de autor

Allan Rocha de Souza ^{1,2}

Luca Schirru ^{2,3}

Miguel Bastos Alvarenga ^{1,2}

doi: 10.1590/0102-311XPT169023

O uso de sistemas informáticos para auxiliar na coleta, organização e análises de grandes volumes de dados e textos e, neste contexto, as técnicas de mineração de textos e dados (*text and data mining* – TDM) ¹ são centrais às pesquisas contemporâneas intensivas em dados, como aquelas conduzidas no combate à pandemia de COVID-19 ^{2,3,4,5,6,7,8}. A par de questões éticas referentes à apropriação dos dados de pesquisa ⁸, os direitos autorais impõem desafios específicos ao uso de TDM, tanto em razão dos objetos sobre os quais incide a proteção, como por sua extensão temporal, amplo escopo de direitos atribuídos e escassez de limitações expressas. Em razão de sua importância para a pesquisa, nos concentramos aqui justamente nestes obstáculos e os movimentos de regulação em curso.

Os direitos autorais não protegem fatos, informações ou dados, e nem mesmo o conteúdo de uma obra, pois seu objeto de proteção é a forma literária ou artística em que são expressos, comunicados. No entanto, quando estes elementos são agrupados, organizados ou sistematizados em uma base ou banco de dados que seja minimamente original com relação à seleção, organização ou disposição de seu conteúdo, este material passa a ter seu acesso e utilização controlado pelo titular deste conjunto, então protegido por direitos autorais ^{9,10,11}.

Ademais, a proteção jurídica por direitos autorais das bases e bancos de dados é apenas uma dentre as camadas de controle sobre o acesso e uso a dados e informações ¹¹. Há uma segunda camada, composta por restrições tecnológicas ao acesso ou uso de bases de dados ¹², e, ao menos, ainda, uma terceira, sob a forma de proibições legais de burla desses mecanismos, mesmo quando não há violação direta de qualquer direito incidente sobre o material ¹³. Isto impacta diretamente atividades essenciais à pesquisa, como verificação, reprodutibilidade e comunicação dos resultados, ao permitir ou limitar quem pode fazer que tipo de pesquisa, utilizando que tipo de material e em que condições ^{14,15,16}. Isto se sobressai na TDM, que necessita da maior disponibilidade de fontes possível – as quais nem sempre existem em acesso aberto. Assim, restringir-se a esse tipo de fonte nesse tipo de método pode levar ao exacerbamento de vieses nos resultados ¹⁵.

Há em curso uma demanda por reformas normativas no âmbito global, diante da extensão normativa, tecnológica e contratual dos poderes de controle pelos titulares, que impõe desafios jurídicos a inúmeras atividades de pesquisas intensivas em dados ^{17,18,19}, o que é especialmente sentido nos países do Sul Global ^{19,20,21}. Diante disso, diversos sistemas jurídicos iniciaram, na última década, reformas em suas legislações de direitos autorais para incluir normas visando assegurar a legalidade da TDM ²², a exemplo da União Europeia ^{23,24}, e de países como Japão ^{25,26} e Singapura ²⁷.

¹ Programa de Pós-graduação em Políticas Públicas, Estratégias e Desenvolvimento, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

² Instituto Brasileiro de Direitos Autorais, Rio de Janeiro, Brasil.

³ Centre for IT & IP Law (KU Leuven), Leuven, Belgium.

Correspondência

M. B. Alvarenga
Rua do Humaitá 282, apto. 1306, bloco 2, Rio de Janeiro, RJ 22261-004, Brasil.
miguel.alvarenga@pped.ie.ufrj.br



No Brasil, há o reconhecimento dos potenciais desafios e oportunidades quanto ao uso de tecnologias intensivas em dados nos textos da Estratégia Brasileira de Inteligência Artificial (EBIA) ²⁸ e na Estratégia Nacional de Propriedade Intelectual (ENPI) ²⁹. Contudo, até agora, a discussão mais aprofundada sobre TDM ocorreu no âmbito dos debates na Comissão de Juristas de Inteligência Artificial (CJUSBIA), responsável por subsidiar a elaboração de substitutivo do *Projeto de Lei* (PL) sobre Inteligência Artificial ³⁰.

O relatório final da CJUSBIA, convertido agora no *PL nº 2.338/2023* do Senado Federal, propõe uma definição de mineração de textos e dados para fins normativos, além de uma limitação (ou exceção) aos direitos autorais do titular de bancos e bases de dados, bem como de outras obras envolvidas em tal processo ³¹. A limitação proposta no art. 42 do *PL nº 2.338/2023* é bastante clara e delimitada em seu escopo: apenas autoriza o acesso e uso por entidades cuja missão é conectada ao interesse público – por exemplo, instituições de pesquisa e de jornalismo – e na extensão necessária para os objetivos pretendidos. A limitação também deixa claro que o uso ali referenciado é aquele que, no que tange a obras protegidas, não implique na criação de produtos concorrentes ou que venham, de alguma forma, impactar na exploração normal das obras utilizadas, por exemplo ³¹.

Alguns defendem que sequer seria preciso uma limitação para mineração de textos e dados ^{24,32}, por ser um uso não-expressivo ^{33,34}; ou, porque o que se extrai das obras – padrões e elementos factuais, por exemplo – não integraria o escopo de proteção dos direitos autorais ^{24,32,34}. Contudo, múltiplos projetos já recorreram ao uso direto de conteúdo protegido por direitos autorais e causaram disputas judiciais, mesmo quando o material havia sido acessado legalmente ³³. Assim, uma previsão normativa expressa é de extrema importância para garantir segurança jurídica às organizações e instituições de pesquisa, jornalismo, museus, arquivos, bibliotecas e também de seus servidores, na satisfação de suas funções e missões institucionais no contexto contemporâneo da pesquisa, inovação e desenvolvimento tecnológico ³⁵.

Mais amplamente, a discussão sobre a legitimidade e importância da TDM para fins de pesquisa tem sido acompanhada de um debate mais amplo sobre o reconhecimento, contornos e efetividade de um direito à pesquisa enquanto garantia fundamental. Regularmente atrelado nos textos normativos, seja ao direito à educação e de acesso ao conhecimento (art. 27 da *Declaração Universal dos Direitos Humanos* e arts. 206, II e 208, V da *Constituição Federal*) ^{36,37}, ou, mais amplamente, à ciência e inovação (art. 218 e seguintes da *Constituição Federal*) ³⁷, suas particularidades demandam independência e autonomia própria, e a elaboração de sua estrutura legal, funções normativas e efeitos jurídicos próprios ainda está em construção ³⁸.

No entanto, a concreção e realização deste direito não têm acompanhado sua relevância social. A recente pandemia expôs a centralidade e a importância da pesquisa e da ciência no funcionamento da sociedade. Os anos recentes nos explicitaram a vulnerabilidade política e social a que as atividades de pesquisa estão sujeitas. Reconhecer a legitimidade da mineração de textos e dados – especialmente, mas não só para fins de pesquisa – é um passo importante e necessário, embora não suficiente, na elaboração e consolidação do direito à pesquisa, que entendemos como axial para fomentar a inovação, o desenvolvimento econômico e autonomia tecnológica do país.

Colaboradores

A. R. Souza contribuiu na concepção do estudo, e análise dos dados, redação e revisão do texto; e aprovou a versão final. L. Schirru contribuiu na concepção do estudo, e análise dos dados, redação e revisão do texto; e aprovou a versão final. M. B. Alvarenga contribuiu na concepção do estudo e análise dos dados, redação e revisão do texto; e aprovou da versão final.

Informações adicionais

ORCID: Allan Rocha de Souza (0000-0002-6549-0085); Luca Schirru (0000-0002-4706-3776); Miguel Bastos Alvarenga (0000-0002-4079-7785).

Referências

- Han J, Pei J, Kamber M. Data mining: concept and techniques. 3rd Ed. Waltham: Morgan Kaufmann; 2012.
- Agrela L. Inteligência artificial previu epidemia do coronavírus da China. Exame 2020; 28 jan. <https://exame.com/tecnologia/inteligencia-artificial-previu-epidemia-do-coronavirus-da-china/>.
- Stieg C. How this Canadian start-up spotted coronavirus before everyone else knew about it. CNBC Make it 2020; 3 mar. <https://www.cnbc.com/2020/03/03/bluedot-used-artificial-intelligence-to-predict-coronavirus-spread.html>.
- Souza AR, Schirru L, Alvarenga MB. Direitos autorais e mineração de dados e textos no combate à COVID-19 no Brasil. Liinc em Revista 2020; 16:e5536.
- You J, Expert P, Costelloe C. Using text mining to track outbreak trends in global surveillance of emerging diseases: ProMED-mail. J R Stat Soc Ser A Stat Soc 2021; 184:1245-59.
- Safdari R, Rezayi S, Saeedi S, Tanhapour M, Gholamzadeh M. Using data mining techniques to fight and control epidemics: a scoping review. Health Technol (Berl) 2021; 11:759-71.
- Pedroso MM, Lima JC, Assef Neto VB. Ciência de dados aplicada ao arca: disponibilização de ferramentas para recuperação da informação no repositório institucional da Fundação Oswaldo Cruz. RECIIS 2017; 11 Suppl. <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1417/pdf1417>.
- Chiruvella V, Guddati AK. Ethical issues in patient data ownership. Interact J Med Res 2021; 10:e22269.
- Brasil. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Diário Oficial da União 1998; 20 fev.
- Organização Mundial do Comércio. Acordo sobre Aspectos dos Direitos de Propriedade Intelectual Relacionados ao Comércio (TRIPS). https://www.wto.org/english/docs_e/legal_e/27-trips.pdf (accessed on 10/Jan/2024).
- Derclaye E. The legal protection of databases: a comparative analysis. Cheltenham: Edward Elgar; 2008.
- Toth AK. Algorithmic copyright enforcement and al: issues and potential solutions, through the lens of text and data mining. Masaryk University Journal of Law and Technology 2019; 13:361-87.
- Brown K. Digital rights management: trafficking in technology that can be used to circumvent the intellectual property clause. Houston Law Rev 2003; 803:803-36.
- Reichman JK, Okediji RL. When copyright law and science collide: empowering digitally integrated research methods on a global scale. Minn Law Rev 2012; 96:1362-480.
- Levendowski A. How copyright law can fix artificial intelligence's implicit bias problem. Washington Law Review 2018; 93:579-630.
- Ducato R, Strowel A. Limitations to text and data mining and consumer empowerment: making the case for a right to machine legibility. CRIDES Working Paper Series 2018; 31 oct. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3278901 (accessed on 30/Jun/2023).
- Caspers M, Guibault L. A right to 'read' for machines: assessing a black-box analysis exception for data mining. Proceedings of the Association for Information Science and Technology 2016; 53:1-5.
- Flynn S, Butler B, Carroll M, Cohen-Sasson O, Craig C, Guibault L, et al. Legal reform to enhance global text and data mining research: outdated copyright laws around the world hinder research. Science 2022; 378:951-3.

19. Souza AR. COVID-19, Text and data mining and copyright: the Brazilian case. In: World Intellectual Property Organization; World Trade Organization, editors. WIPO-WTO Colloquium Papers. v. 11. s/l: World Intellectual Property Organization/World Trade Organization; 2020. p. 1-14.
20. Bertón M. Text and data mining exception in South America: a way to foster AI development in the Region. *GRUR International* 2021; 70:1145-57.
21. Flynn S, Schirru L, Palmedo M, Izquierdo A. Research exceptions in comparative copyright. 2022. <https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1077&context=research> (accessed on 10/Jan/2024).
22. Palmedo M, Alvarenga M, Imran M, Le D, Schirru L. Measuring change in copyright exceptions for text and data mining. <https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=1100&context=research> (accessed on 10/Jan/2024).
23. European Union. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act). <http://data.europa.eu/eli/reg/2023/2854/oj> (accessed on 12/Jan/2024).
24. Margoni T, Kretschmer M. A deeper look into the EU text and data mining exceptions: harmonisation, data ownership, and the future of technology. *GRUR International* 2022; 71:685.
25. World Intellectual Property Organization. Copyright Act (Act No. 48 of May 6, 1970, as amended up to Act No. 72 of July 13, 2018). <https://www.wipo.int/wipolex/en/text/504411> (accessed on 08/Apr/2024).
26. Ueno T. The flexible copyright exception for 'non-enjoyment' purposes - recent Amendment in Japan and its implication. *GRUR International* 2021; 70:145-52.
27. World Intellectual Property Organization. Copyright Act 2021 (Revised edition 2020, Act No. 22 of 2021). <https://www.wipo.int/wipolex/en/text/584840> (accessed on 30/Jun/2023).
28. Ministro de Estado da Ciência, Tecnologia e Inovações. Portaria GM nº 4.617, de 6 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. *Diário Oficial da União* 2021; 9 apr.
29. Brasil. Decreto nº 10.886, de 7 de dezembro de 2021. Institui a Estratégia Nacional de Propriedade Intelectual. *Diário Oficial da União* 2021; 8 dec.
30. Schirru L, Souza AR, Chamas C. Building a text and data mining limitation: the Brazilian case. *GRUR International* 2024; 73:217-22.
31. Brasil. Projeto de Lei nº 2.338/23. Dispõe sobre o uso da Inteligência Artificial. https://legis.senado.leg.br/sdleg-getter/documento?dm=9347593&ts=1684441712901&disposition=inline&_gl=1*9y8waw*_ga*NzYwNzg2OTY4LjE2ODgxMzg3NTQ.*_ga_CW3ZH25XMK*MTY4ODE0MDg1NC4yLjAuMTY4ODE0MDg2Ni4wLjAuMA (accessed on 30/Jun/2023).
32. Senftleben M. Compliance of national TDM rules with International Copyright Law: an overrated nonissue? *International Review of Intellectual Property and Competition Law* 2022; 53:1477-505.
33. Carroll MW. Copyright and the progress of science: why text and data mining is lawful. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3531231 (accessed on 30/Jun/2023).
34. Sag M. The new legal landscape for text mining and machine learning. *Journal of the Copyright Society of the USA* 2019; 66:1-34.
35. Alvarenga MB. Mineração de dados, Big Data e direitos autorais no Brasil [Doctoral Dissertation]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2019.
36. Fundo das Nações Unidas para a Infância. Declaração Universal dos Direitos Humanos (1948). <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos> (accessed on 25/Jan/2024).
37. Brasil. Constituição Federal da República Federativa do Brasil, 05 de outubro de 1988. *Diário Oficial da União* 1988; 5 oct.
38. Geiger C, Jütte BJ. The Right to Research as Guarantor for Sustainability, Innovation and Justice in EU Copyright Law (June 19, 2022). In: Pihljarinne T, Mähönen J, Upreti P, editors. Rethinking the role of intellectual property rights in the post pandemic world: an integrated framework of sustainability, innovation and global justice. Cheltenham/Northampton: Edward Elgar; 2023. p. 138-69.

Recebido em 06/Set/2023

Versão final reapresentada em 29/Jan/2024

Aprovado em 15/Fev/2024