Joint PIJIP/TLS Research Paper Series

10-27-2023

Copyright and Generative AI: Insights from the Project on the Right to Research

Sean M. Fiil-Flynn
sflynn@wcl.american.edu

Recommended Citation

COPYRIGHT AND GENERATIVE AI: INSIGHTS FROM THE PROJECT ON THE RIGHT TO RESEARCH

Sean M. Fiil-Flynn

ABSTRACT

With the widespread launch of "generative AI" technology capable of creating music, art and text that could potentially substitute for the work of human creators, policy makers around the world are highly focused on how and whether copyright law should change in response. This paper summarizes results from studies that have been commissioned as part of the Project on the Right to Research in International Copyright coordinated by American University's Program on Information Justice and Intellectual Property. That project focuses on uses of copyrighted works in text and data mining technology by scientific researchers. The same text and data mining research methods are used as critical inputs in generative AI applications. Accordingly, policymakers need to mindful of the overlapping uses less policy responses to one problem (promoting creativity) cause another (restriction of research methods).

FINDINGS OF THE PROJECT ON THE RIGHT TO RESEARCH IN INTERNATIONAL COPYRIGHT

The following propositions are supported by the published research undertaken as part of the Project on the Right to Research in International Copyright.

I. TEXT AND DATA MINING AND GENERATIVE AI

One necessary input to most AI tools, including but not limited to "generative AI" large language models, is text and data mining research (TDM) - which uses computers to analyze digitized information.[1]

TDM is used to create predictive models used in Generative AI, but is also used in critical academic research in the physical sciences, social sciences and humanities.[2]

II. THE FUNDAMENTAL RIGHT TO RESEARCH

Academic and scientific research by humans, including through technological methods such as TDM, implicates fundamental rights to receive and impart information and to benefit from culture and science -- which may be referred to as the "right to research."[3]

TDM, AI and machine learning research are taking place in developing countries where researchers are often reliant on free-to-use text and data sets.[4]

III. COPYRIGHT BARRIERS TO TEXT AND DATAMINING RESEARCH

A patchwork of copyright laws across jurisdictions limits where and how TDM research, and therefore the construction of various AI tools and applications, can occur.[5]

---

[1] Sean Flynn et al., *Implementing User Rights for Research in the Field of Artificial Intelligence* (American Univ. Wash. College of Law Research Paper No. 2020-12, 2020).

[2] Sean Flynn & Lokesh Vyas, *Examples of Text and Data Mining Research Using Copyrighted Materials*, Mar. 6, 2023, https://copyrightblog.kluweriplaw.com/2023/03/06/examples-of-text-and-data-mining-research-using-copyrighted-materials/

[3] *See* Klaus Beiter, *Reforming Copyright or Toward Another Science? A More Human Rights-Oriented Approach Under the REBSPA in Constructing a "Right to Research" for Scholarly Publishing*, PIJIP/TLS Research Paper Series no. 7 (2022); Desmond O. Oriakhogba, *The Right to Research in Africa: Making African Copyright Whole*, PIJIP/TLS Research Paper Series no. 78 (2022); Marcela Palacio Puerta, *Transitional Justice, Truth, and Copyright: The Case of Colombia*, PIJIP/TLS Research Paper Series no. 95 (2023).

[4] Pijipvideo, *Prof Vukosi Marivate: NLP and TDM in Africa. Right to Research in Int'l Copyright Seminar 11.2*, YouTube (March 22, 2023); Flynn, *supra* note 2.

[5] Sean Flynn et al., *Legal Reform to Enhance Global Text and Data Mining Research*, Science (Dec. 1, 2022), https://www.science.org/doi/10.1126/science.add6124; *Is it legal? New database with analysis of hypothetical cases on exceptions and limitations to copyright in Latin America*, Latin American Civil Society Alliance, https://accesoalconocimiento.lat/2023/02/28/es-legal-nueva-base-de-datos-con-analisis-de-

The conflicting state of national laws creates particular barriers to cross-border research projects that need to share materials between jurisdictions.[6]

Providing copyright exceptions or other clarifications of the law to permit TDM is associated with increased publication of scientific research in the countries that make the change.[7]

Exclusivity and licensing market models to regulate access and use of text and data for research, including for TDM and AI applications, chills useful research and economic activity.[8]

The chilling effect of market-based models is likely to be proportionally higher for resource dependent researchers, including those in developing countries, because information monopolies tend to price information at global uniform(ly high) levels.[9]

The chilling effect of restrictive copyright also affects the development of AI industries, including especially in developing countries.[10]

Providing open research exceptions in copyright that apply to "non-expressive" uses (i.e. that do not themselves share copies with the public) of all works by all users is justified by historic and theoretical copyright analysis in common law and civil law countries.[11]

Claims about the potential harms to creators and content owners from

---

casos-hipoteticos-sobre-excepciones-y-limitaciones-al-derecho-de-autor-en-america-latina/; desarrolloa2k, *Report on Copyright Flexiformities in Latin America*, May 17, 2021, https://accesoalconocimiento.lat/2021/05/17/informe-sobre-flexibilidades-al-derecho-de-autor-en-america-latina/

[6] Naama Daniel, *Lost in Transit: How Enforcement of Foreign Copyright Judgements Undermines the Right to Research*, Joint PIJIP/TLS Research Paper Series. 85 (2023); Sean Flynn et al., *Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action*, Joint PIJIP/TLS Research Paper Series. 48 (2020).

[7] Pijipvideo, *Empirical Study Pt2: Impact of Research Exceptions on Scientific Output-Joan-Josep Vallbé, - May 23*, YouTube (July 24, 2023), https://www.youtube.com/watch?v=2bs_e7QRDHo&list=PLuk2SmOxN5RI1z40tC6qDxV6uQdq-kqLq&index=4; Michael Palmedo, *The Impact of Copyright Exceptions for Researchers on Scholarly Output*, 2 Efil Journal 114 (2017).

[8] Patricia Aufderheide, *The Chilling Effect of Copyright Permissions on Academic Research: The Case of Communication Researchers*, Joint PIJIP/TLS Research Paper Series. 49 (2020); Anthony D. Rosborough & Aaron Perzanowski, *Repair as Research: How Copyright Impedes Learning About Devices*, PIJIP/TLS Research Paper Series no. 101 (2023).

[9] Sean Flynn et al., *An Economic Justification for Open Access to Essential Medicine Patents in Developing Countries*, 37 J.L. MED. & ETHICS 184 (2009).

[10] *Report on Artificial Intelligence and Copyright Policies in Latin America*, Latin American Civil Society Alliance, https://accesoalconocimiento.lat/2023/03/03/informe-sobre-politicas-de-inteligencia-artificial-y-derechos-de-autor-en-america-latina/

[11] Professor Matthew Sag, *Fairness and Fair Use in Generative AI*, 12th Annual Peter A. Jaszi Distinguished Lecture on Intellectual Property (Sept. 28, 2023); Christophe Geiger, *"Fair Use" through Fundamental Rights in Europe: When Freedom of Artistic Expression allows Creative Appropriations and Opens up Statutory Copyright Limitations*, PIJIP/TLS Research Paper Series no. 63 (2020).

more open copyright exceptions are not supported by empirical research.[12]

There is a substantial and growing trend in copyright reform toward the adoption of more open copyright exceptions, including most recently for research uses.[13]

Adoption of more open research exceptions is fully compliant with international copyright law.[14]

The right to research can and should be articulated and promoted in international copyright law and programs, including in the exceptions of any final broadcast treaty, in an instrument on limitations and exceptions, and in WIPO projects and budgets.[15]

---

[12] Niva Elkin-Koren & Neil Weinstock Netanel, *Transplanting Fair Use Across the Globe: A Case Study Testing the Credibility of U.S. Opposition*, Joint PIJIP/TLS Research Paper Series. 50 (2020); Sean Flynn & Michael Palmedo, *The User Rights Database: Measuring the Impact of Opening Copyright Exceptions*, Joint PIJIP/TLS Research Paper Series no. 2018-01 (2018).

[13] Sean Flynn & Michael Palmedo, *The User Rights Database: Measuring the Impact of Opening Copyright Exceptions*, Joint PIJIP/TLS Research Paper Series no. 2018-01 (2018).

[14] Michael Palmedo et al., *Measuring Change in Copyright Exceptions for Text and Data Mining*, PIJIP/TLS Research Paper Series no. 98 (2023)

[15] Sean Flynn et al., *Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action*, 42 European Intellectual Property Review 393 (2020); Bernt Hugenholtz, *The WIPO Broadcasting Treaty: Comments on the Second Revised Draft*, Joint PIJIP/TLS Research Paper Series. 84 (2023); World Intellectual Property Organization, *Proposal by African Group for a Draft Work Program on Exceptions and Limitations*, March 20, 2023, https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=603511; World Intellectual Property Organization, *Revised Pilot Project on Text and Data Mining (TDM) to Support Research and Innovation in Universities and Other Research-Oriented Institutions in Africa – Proposal Submitted by the African Group*, April 28, 2023, https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=607691.