

Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for Action at International Level

Kluwer Copyright Blog
April 21, 2020

Sean Flynn (Washington College of Law), Christophe Geiger (University of Strasbourg – CEIPI), and João Pedro Quintais (Institute for Information Law (IVIR))

Please refer to this post as: Sean Flynn, Christophe Geiger, and João Pedro Quintais, 'Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for Action at International Level', Kluwer Copyright Blog, April 21, 2020, <http://copyrightblog.kluweriplaw.com/2020/04/21/implementing-user-rights-for-research-in-the-field-of-artificial-intelligence-a-call-for-action-at-international-level/>

Last year, before the onset of a global pandemic highlighted the critical and urgent need for technology-enabled scientific research, the World Intellectual Property Organization (WIPO) launched an inquiry into issues at the intersection of intellectual property (IP) and artificial intelligence (AI). We and other members of the [Global Expert Network on Copyright User Rights](#) contributed comments to that inquiry, with a focus on the application of copyright to the use of text and data mining technology. This blog, and [the article on which it is based](#), describe some of the most salient points of our submission and conclude by stressing the need for international leadership on this important topic.



Copyright and text and data mining research

“Text and data mining” (“TDM”) is often a necessary step to train machine learning systems or processes. TDM includes any application of a computational process to materials to derive data from or about those works. TDM can be used to discover new facts, such as correlations, patterns and links between information points in the database. “Machine learning” applies additional analysis and processes to information often gleaned from TDM to enable machines to dynamically “learn” new tasks for which they were not specifically programmed. The term “artificial intelligence” or “AI” is often used as an umbrella term to describe a number of technologies or systems, including what we define as “machine learning” or an advanced application of it (e.g. deep learning), as well as evolutionary algorithms and rules-based systems.

Many of the most useful TDM and AI projects involve the use of copyright-protected works. The [BlueDot project that discovered the novel coronavirus outbreak](#), for example, analyzed “a variety of information sources, including chomping through 100,000 news reports in 65 languages a day” to recognize patterns between health outbreaks and travel.

Engaging in TDM often entails making both temporary and permanent reproductions of copyrighted works. Temporary reproductions are made any time a researcher makes a query of a database. These copies may be fleeting and, as such, could fall within limitations and exceptions (L&Es) for the making of transitory copies of works for the purpose of facilitating a technological process, which are provided for in many laws. TDM research also requires the making of more permanent copies to construct the database of works to be mined. Additional reproductions and communications of the database itself may be necessary to permit other researchers to use it and to test for accuracy, replicability and transparency. The question for global copyright rules is therefore whether any of these uses of works fall within the exclusive rights of copyright holders, for which a license must be obtained. At policy level, the question is how to ensure that such vital uses for innovation and research are facilitated and not hindered by the copyright system.

Copyright law provides protection for the material interests of authors through rights to exclude certain uses of their works, including their reproduction. At the same time, one of the universally accepted axioms of copyright law is that exclusivity should apply only to original expression, not to facts, ideas, procedures or methods of operation. It is also universally accepted that copyright contains free spaces to ensure follow-on creativity and to secure important fundamental rights and the public interest, in particular allowing research to be undertaken using protected material.

TDM reproductions do not compromise the core interest of exclusive rights, which is to prohibit unauthorized reproductions that can substitute for the work of the author. This is because TDM makes only “non-expressive” uses of works. It could even be argued that these incidental reproductions are outside the scope of exclusive rights. Also, as has been underlined by several scholars, mere reading does (and should) not involve a copyright relevant action, and neither should “the act of reading a work into a computer’s random access memory.” Denying the ability to make reproductions of works needed to undertake TDM would deny the possibility to read and to access to the very ideas, facts and data at the root of these works, thereby limiting the enjoyment of what we refer to as a “right to research.”

The right to research

Rights to conduct and receive or access research have a strong fundamental rights justification, in particular with regard to freedom of information and the public’s right to information. In part to serve these fundamental interests, privileged uses to conduct research with materials protected by IP law are quite common. These protections of research activities can be found in restrictions on the scope of exclusive rights or through the provision of L&Es. In addition to promoting research through the exclusion of facts from the scope of protection, copyright laws frequently contain L&Es for uses of protected works for “research” or “private study.”

For some TDM processes – such as making a query of an existing database – the exclusion of facts and ideas from copyright protection may be sufficient to authorize the activity. But other actions, such as the creation of a database of reproductions for the mining process, appear to require explicit authorization. To provide such authorization, the laws of a growing number of jurisdictions around the world are recognizing L&Es to exclusive rights for “text and data mining,” “information analysis”, “computational analysis,” or similar activities or purposes. In its latest copyright revision, the EU has recently introduced specific limitations and exceptions for this purpose, unfortunately with several restrictions, thus leaving significant uncertainties on the legality of many TDM activities (for a comment on these provisions, [see here](#)).

A Role for WIPO

To promote the greatest possible use of technology in scientific research, we encourage WIPO to take the lead in promoting TDM exceptions in every country, including through the development of an international agreement on cross border uses of TDM databases.

In its technical assistance activity, we suggest that WIPO promotes the use of “open” exceptions for TDM research. These are exceptions which, like in Japanese and other laws, authorize fair practices for research purposes that apply to all kinds of copyright-protected works, all kinds of exclusive rights, and to acts by all users. In this regard, the TDM exceptions in the 2019 EU Copyright in the Digital Single Market (CDSM) Directive are not best practice.

In its norm setting work, we encourage WIPO to work on an international instrument to facilitate cross border sharing of TDM tools and databases. The failure to use open models for TDM exceptions in some countries, and the absence of L&Es in many others, has created ambiguities about whether researchers can collaborate on cross border TDM projects. This is true even for collaborations between countries where TDM activities are legal in each.

Imagine a researcher in the EU, where making a TDM database would be lawful under the CDSM Directive, collaborating with a researcher in the U.S., where TDM is also lawful under fair use rights. Can the EU researcher transfer a database lawfully made in the EU to the partner researcher in the US? The answer is unclear (at best), because the respective TDM right applies only to the “reproduction” right under copyright, not to the separately protected communication and making available rights.

This is a problem that WIPO is uniquely situated to solve. The [Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired, or Otherwise Print Disabled](#) adopted a novel international rule permitting the cross border exchange of accessible materials lawfully made in any member country. WIPO’s Standing Committee on Copyright and Related Rights could consider a similar norm for the cross border sharing of lawfully produced research materials for TDM.