

WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)

Second Session

**DRAFT ISSUES PAPER ON INTELLECTUAL PROPERTY POLICY AND
ARTIFICIAL INTELLIGENCE**

Submitted by:

Andrés Izquierdo A.

International Copyright Policy Advisor and
Copyright Lawyer, USA - Colombia

Gustavo Palacio Correa

Policy Advisor and Copyright Lawyer,
Colombia

The World Intellectual Property Organization [requests comments](#) on a list of questions concerning the impact of artificial intelligence (AI) on intellectual property (IP).

This comment was prepared by Andrés Izquierdo and Gustavo Palacio, international copyright policy advisors and copyright lawyers.

We only comment here on the copyright related questions in section 13.

13(i). Should the use of the data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright? If not, should an explicit exception be made under copyright law or other relevant laws for the use of such data to train AI applications?

A. This question smartly identifies the current problem. We are trying to decide whether the use of text and data mining (TDM) to feed the machine learning process should be legal or not, if there should be exceptions, or if we should not allow TDM as a general rule of application. However, we considered that before looking for answers outside the current system, we would look for answers within the current legal instruments. Initial questions would be as follows:

1. In which cases does the use of TDM for machine learning constitute copyright violation under the current WIPO treaties (Berne, Trips, WCT, WPPT)
2. Which exceptions could possibly apply for TDM for machine learning under Berne, Trips, WCT, or WPPT?
3. Does the dichotomy idea/expression found implicitly or explicitly on Berne, Trips, WCT, or WPPT be applied as an exception for machine learning?

4. Does the current copyright expression requirement in copyright law can solve the current questions on data for machine learning?
5. Could we use the three-step test for TDM applied to machine learning?

B. We also believe in the need to clearly identify what does machine learning imply and how does it differ from artificial intelligence (AI). Machine learning is based on the idea that machines should be able to learn and adapt through experience, while AI refers to a broader idea where machines can execute tasks applying machine learning, deep learning and other techniques to solve actual problems.

Question: 1. What constitutes machine learning and how does it relate to TDM?

C. Normally machine learning follows two steps: first, the data is collected, organized, in some cases transformed, with the objective of creating a corpus; second, the data is uploaded to the computer in order to generate the machine learning process. These two steps brings about two legal moments: 1. collection of data and 2. upload. Initial questions would be as follows:

1. Is the data gathered without authorization to create a corpus should constitute an infringement of copyright?
2. Does uploading and processing of the corpus for machine learning should constitute a copyright infringement?
3. If there is a violation on copyright in either or both of the previous steps, should an explicit exception be made under copyright law or other relevant laws for the use of such data?

13(ii). If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, what would be the impact on the development of AI and on the free flow of data to improve innovation in AI?

To address this question, it is important to consider the different actors in the market and their interests, such as the corporations in the entertainment business, technology and services industries, governments, users, artists, universities, researchers, libraries, among others. There might also be undiscovered interests, such as deep web AI developers. To be able to identify property the actors involved I would frame the question as follows:

1. Please name the different actors and interests that are negatively or positively affected by the use of text and data mining for the machine learning process.

13(iii). If the use of the data subsisting in copyright works without authorization for machine learning is considered to constitute an infringement of copyright, should an exception be made for at least certain acts for limited purposes, such as the use in non-commercial user-generated works or the use for research?

This question tries to decide if there should be a distinction between commercial and non-commercial use for text and data mining.

The tendencies around the world show a division from the countries that have regulated the topic and the ones that have not. There are too many countries that still remain with no laws or knowledge on the matter, as most of the countries in Latin America, Middle East, or Africa. Countries that have regulated the topic include some of the countries in Europe, as well for the United Kingdom, United States, Japan, and Canada.

From the countries that have regulated the matter, there is a distinction between the ones that have a commercial and non-commercial use exception, and the countries that do not make that distinction. Among the firsts one you can find Germany, France, United Kingdom, and among the second ones, you can find United States, Canada or Japan.

The United States exception is created by judicial interpretation of the fair use clause of the copyright act Sec. 107 of the US Code. In Japan, the restriction is by law. It's important to note that the application of the US case law could be very narrow, currently giving lack of certainty to the parties interested in having a default position on the matter. Also, the relevant US cases (Google Books and HathiTrust) did not address issues arising under laws prohibiting computer hacking, contract law, cross-border copyright issues, or laws prohibiting the circumvention of technological protection measures.

Given the previous, I would suggest a slight change in the framing of the question:

1. Should an exception for the use of unauthorized copyrighted data for machine learning make the distinction between commercial and non commercial use?

Also the exception should consider factors such as: purpose of the work, exclusive rights, transfer and sharing, lawfully accessed contents, cross border rights, contract and Technological protections measures override.

13(iv). If the use of the data subsisting of copyright works without authorization for machine learning is considered to constitute an infringement of copyright, how would existing exceptions for text and data mining interact with such infringement?

It may be best to combine this question with the question above.

Additional proposed questions:

Ethical considerations:

There are many ethical issues that have arisen alongside this technological development. Facial recognition is being used by governments, employee morale is negatively affected when machines take over jobs, AI can be biased (is the AI information fair and neutral?), actors and voices can be cloned (there are many claims that some famous actresses have been cloned in productions rated PG-13 and above), accelerated hacking is a problem, and one of the hottest concerns yet: AI terrorism. AI terrorism can consist of autonomous drones, robotic swarms, remote attacks, or delivery of disease through nanorobots. Let's not forget a very important one:

humanity. It's clear that machines are already affecting our behavior and interactions. So questions about ethical implications about machine learning and AI are also needed:

1. How to limit AI so it will not continue to affect negatively our human condition, human relations, human existence?

Diverse points of view:

1. Should machine learning follow the rules of the traditional copyright systems (Berne, TRIPS, WCT, WPPT)?
2. Should machine learning follow a common law fair use approach (USA)?
3. If we decide to use the fair use approach, should fair use be the policy-making mechanism for machine learning and artificial intelligence, in other words, shall we give the policy making job to the judges?
4. If so, are the judges fit for this type of policy making decisions?
5. When there is free access to data for TDM activities, who are the ones that are benefiting the most of this position?

We also suggest some relevant case law for review:

United States

1. *Authors Guild v. Google*, 804 F.3d 202 (2d Cir. 2015).
2. *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).
3. *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007).
4. *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 933 (2005).
5. *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984).
6. *Ty, Inc. v. Publ'ns Int'l. Ltd.*, 292 F.3d 512, 520 (7th Cir. 2002).
7. *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003).
8. *Harper v. Row Publishers, Inc.*, 471 U.S. 539 (1985).
9. *Baker v. Selden*, 101 U.S. 99 (1879).
10. *Feist Publ'ns, Inc. V. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991).
11. *Warner Bros. Pictures, Inc. v. Columbia Broadcasting Systems, Inc.*, 216 F.2d 945 (1983)
12. *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2d Cir. 1930)
13. *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240 (3d Cir. 1983).
14. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586–87 (1994)
15. *Ticketmaster Corp. v. Tickets.com, Inc.*, No. CV997654HLHVBKX, 2003 WL 21406289, at *1 (C.D. Cal. Mar. 7, 2003).

16. *A.V. ex rel. Vanderhye v. iParadigms, L.L.C.*, 562 F.3d 630 (4th Cir. 2009).

Europe

1. *SAS Institute, Inc v. World Programming Ltd*, European Court of Justice
2. *Hollinrake v. Truswell*, 3 Ch. 420 (Court of Appeals 1894)
3. *Ibcos Computers Ltd. v Barclays Mercantile Highland Finance Ltd*, 1994