Works as Data: Assessing Machine Learning from an EU Copyright Law perspective

Dr. Theodoros Chiou, Phd in Intellectual Property Law, Post-Doc Researcher, Faculty of Law, University of Athens - Lawyer

Abstract

In our days, Artificial Intelligence (AI) is described as "the new electricity", as AI systems that emulate intelligent behavior in terms of computational processes (Georgouli, p. 13), are (or are about to be) put into daily service of human activity, including online activity.

Today's blossom of AI is based mainly on the production and collection of enormous amounts of data and the augmentation of computational power. Moreover, current algorithmic innovation allowed the development of software which enables machines to learn and to make their own decisions and actions, on the basis of a process widely known as Machine Learning (ML) (Georgouli, p. 127 ff).

Within the ML context, Machines are repeatedly trained by means of specifically designed learning algorithms that use as training material a corpus of examples (data sets or training material) and labels that contain the rules that govern the relations between incoming information and produced results (Surden, p. 89 ff). Current state of algorithmic training protocols and Machine Learning techniques reserves the possibility for machines to achieve autonomous decision making, with limited or no human involvement, in a vast number of applications, such as speech recognition, machine translation or algorithmic creation of works of art.

Very often and, especially in the context of algorithmic creations, the training material for intelligent machines is composed by copyrighted works, such as texts, scientific papers, images, musical compositions and other. However, Machine Learning process and training typically involves the realization of one or more reproductions of any work used as training data (Margoni, p. 1). As a consequence, a copyright-related challenge arises: under which conditions the use of protected works as training material within Machine Learning processes is lawful from a copyright law perspective?

The present paper aims at assessing the extent to which the implementation of Machine Learning techniques that use *corpora* of copyrighted works as training data is covered by the monopolistic power of the copyright owner. More precisely, it examines whether such Machine Learning processes require prior authorization from rightholders or whether such use could be considered as lawful, even without prior authorization, due to the application of one or more copyright exceptions.

The analysis will be undertaken on the basis of some examples of ML projects, under the spectrum of European Copyright Law, including the forthcoming Directive on

copyright in the Digital Single Market (DSM Directive)[1]. It will, ultimately, assess whether current EU Copyright Law copes with the new dynamics and meets one of the several challenges set by the Era of the 4th industrial revolution and Web 4.0, as works are not perceived anymore merely as content but mainly as data.

Bibliography

D'ARGENTEUIL MEEÛS, DE FRANCQUEN A. (2014), Study of the Legal Framework of Text and Data Mining (TDM), prepared for the European Commission, p. 28, available at:http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf

GEIGER Ch., FROSIO G., BULAYENKO O., "Opinion of the CEIPI on the European Commission's Proposal to Reform Copyright Limitations and Exceptions in the European Union", Article (PDF Available), SSRN Electronic Journal, January 2017.

GEIGER Ch., FROSIO G., BULAYENKO O., "Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?" (July 5, 2018). 48 *IIC - International Review of Intellectual Property and Competition Law* (2018). Available at SSRN: https://ssrn.com/abstract=3260037

GEIGER Ch., FROSIO G., BULAYENKO O., Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market (2018). in INTELLECTUAL PROPERTY AND DIGITAL TRADE IN THE AGE OF ARTIFICIAL INTELLIGENCE AND BIG DATA 97-111 (Xavier Seuba, Christophe Geiger, and Julien Pénin eds., CEIPI/ICTSD publication series on "Global Perspectives and Challenges for the Intellectual Property System", Issue No. 5, Geneva/ Strasbourg) . Available at SSRN: https://ssrn.com/abstract=3260057

GEIGER Ch., FROSIO G., BULAYENKO O., The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market-Legal Aspects, Study for the European Parliament (JURI Committee), PE 604.941-February 2018, available at: http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf

GEORGOULI K., Artificial Intelligence, An introductory approach, Hellenic Academic Electronic Textbooks, available at: www.kallipos.gr

MARGONI Th., "Artificial Intelligence, Machine Learning and EU Copyright Law: Who Owns AI?", Article (PDF Available) in SSRN Electronic Journal January 2018

QUINTAIS J.-P., "Rethinking Normal Exploitation: Enabling Online Limitations in EU Copyright Law", *AMI*, no 6, 2017, pp. 197-205

[1] Directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), adopted on 26 March 2019.

SURDEN H., “Machine Learning and Law”, *Washington Law Review*, Vol. 89, No. 1, 2014

Keywords: Artificial Intelligence, Machine Learning, Algorithmic creation, Copyright Law, EU acquis, DSM Directive, copyright exceptions and limitations