

Copyright Clearance for Artificial Intelligence Training Data: Challenges and Solutions

Authors : Erva Akin

Abstract : - The use of copyrighted material for machine learning purposes is a challenging issue in the field of artificial intelligence (AI). While machine learning algorithms require large amounts of data to train and improve their accuracy and creativity, the use of copyrighted material without permission from the authors may infringe on their intellectual property rights. In order to overcome copyright legal hurdle against the data sharing, access and re-use of data, the use of copyrighted material for machine learning purposes may be considered permissible under certain circumstances. For example, if the copyright holder has given permission to use the data through a licensing agreement, then the use for machine learning purposes may be lawful. It is also argued that copying for non-expressive purposes that do not involve conveying expressive elements to the public, such as automated data extraction, should not be seen as infringing. The focus of such 'copy-reliant technologies' is on understanding language rules, styles, and syntax and no creative ideas are being used. However, the non-expressive use defense is within the framework of the fair use doctrine, which allows the use of copyrighted material for research or educational purposes. The questions arise because the fair use doctrine is not available in EU law, instead, the InfoSoc Directive provides for a rigid system of exclusive rights with a list of exceptions and limitations. One could only argue that non-expressive uses of copyrighted material for machine learning purposes do not constitute a 'reproduction' in the first place. Nevertheless, the use of machine learning with copyrighted material is difficult because EU copyright law applies to the mere use of the works. Two solutions can be proposed to address the problem of copyright clearance for AI training data. The first is to introduce a broad exception for text and data mining, either mandatorily or for commercial and scientific purposes, or to permit the reproduction of works for non-expressive purposes. The second is that copyright laws should permit the reproduction of works for non-expressive purposes, which opens the door to discussions regarding the transposition of the fair use principle from the US into EU law. Both solutions aim to provide more space for AI developers to operate and encourage greater freedom, which could lead to more rapid innovation in the field. The Data Governance Act presents a significant opportunity to advance these debates. Finally, issues concerning the balance of general public interests and legitimate private interests in machine learning training data must be addressed. In my opinion, it is crucial that robot-creation output should fall into the public domain. Machines depend on human creativity, innovation, and expression. To encourage technological advancement and innovation, freedom of expression and business operation must be prioritised.

Keywords : artificial intelligence, copyright, data governance, machine learning

Conference Title : ICPRIP 2023 : International Conference on Property Rights and Intellectual Property

Conference Location : Tokyo, Japan

Conference Dates : July 17-18, 2023