

A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data ownership, and the future of technology

Thomas Margoni, Research Professor of Intellectual Property Law, CiTiP, KU Leuven.

Martin Kretschmer, Professor of Intellectual Property Law and Director, CREATE, University of Glasgow.

31 January 2022

Abstract submitted for consideration at the 40th ATRIP Annual Congress

Abstract: There is global attention on new data analytic methods. Data scraping, text and data mining, and machine learning, are seen as critical technologies. The ongoing pandemic and connected health crisis have shown once more how crucial is a proper regulatory framework for data able to balance opposing social and economic drivers. The legal issues involved in the regulation of data range from privacy and data protection (such as the GDPR) to proprietary approaches (such as copyright, database rights, or proposed new rights in data themselves).

This paper focusses on one specific intervention, the introduction of two exceptions for text and data mining in the Directive on Copyright in the Digital Single Market (CDSM). Art. 3 is a mandatory exception for text and data mining (TDM) for the purposes of scientific research by research and cultural institutions; Art. 4 permits text and data mining by anyone but with rightsholders able to “contract-out”.

We trace the context of using the lever of copyright law to enable emerging technologies and support innovation. Within the EU copyright intervention, elements that may underpin a transparent legal framework for AI are identified, such as the possibility of retention of (permanent) copies for further verification. On the other hand, we identify several pitfalls, including an excessively broad definition of TDM which makes the entire field of data-driven AI development dependent on an exception. We analyse the implications of limiting the scope of the exceptions to the right of reproduction, we argue that the limitation of Art. 3 to certain beneficiaries remains problematic, and suggest that the requirement of lawful access is difficult to operationalize.

In conclusion, we recommend that there should be no need for a TDM exception for the act of extracting unprotected informational value from protected works. The EU’s CDSM provisions paradoxically may favour the development of biased AI systems due to price and accessibility conditions for accessing training data that offer the wrong incentives. We also identify some old and new areas of the EU *acquis* which will play a crucial role in the future relationship of EU copyright law with technological development.