

## European Alliance for Research Excellence

Answer to WIPO's draft issues paper on intellectual property policy and artificial intelligence  
([WIPO/IP/AI/2/GE/20/1](#))

The [European Alliance for Research Excellence](#) (EARE) is a coalition of companies and research organisations formed in 2017 that are committed to the future of innovation and R&D in Europe. The coalition is supported by [BSA | The Software Alliance](#), [Allied for Startups](#), [Research Libraries UK](#), [SCONUL](#) (Society of College, National and University Libraries) and [UCL](#) Library (University College London), and has been advocating for copyright rules in Europe that enable a fair and effective use of Text and Data Mining (TDM), to ensure Europe's competitiveness and future prosperity.

Since 2017, we have been supporting a broad, easy-to-understand TDM exception in Europe. TDM generates **actionable intelligence from data sets** that were once too large and too volatile to analyse, and is driving revolutionary advances in data analytics, machine learning, and artificial intelligence that are helping address some of society's most pressing challenges.

We warmly welcome the opportunity to provide comments on WIPO's draft issues paper on intellectual property policy and artificial intelligence ([WIPO/IP/AI/2/GE/20/1](#)). We believe that the **key issues related to TDM are already correctly identified** in the draft issues paper, namely through points b) Copyright/Issue 7) Infringement and Exceptions and c) Data 10) Further Rights in Relation to Data.

However, EARE members wish to highlight issues and questions that are highly relevant for the TDM discussion, while at the same time **suggesting additional questions to be considered to ensure the appropriate IP framework is put in place** to take full advantage of the opportunities provided by the data economy and train artificial intelligence applications.

### **b) Copyright/Issue 7) Infringement and Exceptions**

#### **Question 13.(i)**

We strongly believe that the right exceptions to copyright are required to support the development and training of AI applications which offer vast potential for economic growth and increased competitiveness globally. However, EARE members are **concerned that the formulation of the questions under point 13 assumes that "the use of data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright"**. In this regard, it is important to note that TDM and machine learning are **not about enabling access to copyrighted material for free**, but on the contrary, they are about understanding the works accessed legally to identify patterns, facts, and correlations locked within these works.

TDM and machine learning may require automated, incidental storage of lawfully accessed copyrighted works to access non-copyrighable information. However, the results of TDM, such as

knowledge extraction and pattern recognition, do not result in a copy or substantial taking of the material from which the data is used to train the AI application.

As such, the output from performing text and data mining is not a copyright infringement. The process and results of TDM do not imply the underlying expressive value of the copyrighted work, and do not interfere with economic value or business models associated with publications. In addition, international copyright frameworks, such as TRIPS, recognise that facts and data should not be covered by copyright.

We would therefore **recommend WIPO's Secretariat to also consider adding the following questions** to its revised Issues Paper:

- *Q: Given the importance of access to data to develop AI, how should the international copyright system ensure that copyright subject matter does not extend to "ideas, procedures, methods of operation or mathematical concepts as such", in accordance with TRIPS Article 9(2)?*
- *Q: To allow for reproduction and verification of the results of a TDM research, databases including incidental copies of lawfully accessed copyrighted works need to be built and preserved. How can IP policy ensure such databases can effectively be preserved and re-used?*

EARE members believe that **this aspect is a crucial dimension** of the debate on the development of AI.

Indeed, Text and Data Mining is a building block for both machine and deep learning, and machine learning plays a foundational role in the development of AI. Analysing data using TDM is essential to enable machines to learn and AI algorithms to better recognize, understand, adjust, and respond. Without the ability of computers to access and analyse very large amounts of data, employ cognitive technologies to allow the learning of patterns, AI would simply not be possible.

### **Question 13.(iii)**

We believe that copyright laws need to reflect the realities of 21st century research, where the growing use of big data and artificial intelligence tools in research and innovation are necessary to achieve breakthroughs. As such, **Japan, the US, Canada, China and most recently the EU have all adopted broad mandatory exceptions to their copyright laws to remove barriers to text and data mining and enable machine learning**, with the objective to take full advantage of the opportunities provided by the data economy and train AI applications.

In a modern digital economy, public interest researchers are not alone in dealing with vast amounts of data which they need to make sense of. Researchers across all spectrums, in universities, in businesses, in startups, in public-private collaborations need the ability to analyse and understand the data they have legal access to. An exception limited to non-commercial or research purposes only would mean that private companies and spinoffs of public interest research projects are left in a legal grey zone that hampers what they can do with their research and innovation.

Drawing an artificial line between the entities that engage in "research" eliminates the use of data mining and machine learning in any research that could produce innovation which might get

commercialised. This distinction would directly impact funding for research projects of all types and threatens to choke off valuable collaboration with private industry.

Given the fact that jurisdictions have already taken a wider approach to the exception, **we think that question 13.iii) as reproduced above should be reworded so as to allow for responses to look beyond the commercial/non-commercial dichotomy**, which we believe is artificial considering the realities of today's research and innovation:

- *Q: How could a broad and encompassing copyright exception encourage the uptake of Artificial Intelligence while ensuring copyrighted works are properly protected from infringement?*

#### **Question 13.(iv)**

Existing exceptions for text and data mining permit reproductions of lawfully accessed works for the purposes of computational analysis without authorization, in order to facilitate innovation and in recognition that such uses do not harm legitimate interests of copyright owners. In Europe, such exceptions permit copyright and database owners to reserve the use of their works for text and data mining by commercial entities, but **require owners to take affirmative, machine readable steps to avoid application of the exception**.

In other countries, exceptions for text and data mining are absolute, such as those implemented in Japan, and Singapore, which also permit related reproductions such as storage of copyrighted works to access and validate computational analysis results.

In countries where a use could be deemed to be infringing absent authorization or an exception, the interaction with such exceptions should be to enable text and data mining unless it can be demonstrated that such use clearly interferes with and undermines the legitimate expectations of copyright owners for exploitation of such works. In the case of text and data mining that does not result in the abstraction of recognizable copyrightable subject matter, no infringement should be deemed to occur, notwithstanding the need to make temporary copies of such works in order to extract and analyse the data contained within.

We would **recommend WIPO's Secretariat to reword question 13.iv)** to clarify the actions that content owners can take to reserve their rights.

#### **Question 13.(v)**

We believe that the introduction of licenses for the purpose of machine learning would be problematic for a variety of reasons:

1. Machine learning relies on text and data mining, which works by crawling thousands of different digital sources. As long as the TDM user has legal access to a copyright-protected work, either through a license or because they are freely available, they should not have to acquire an additional license to mine that content for the purposes of machine learning. Text and data mining is an automated way to read content one already has access to and should not be subject to additional licenses.

2. Experience has shown that licenses for TDM have been granted only in very narrow and specific fields across a relatively small number of publications, not nearly broad enough to support the type of machine learning that is needed to reap the full benefits of Artificial Intelligence. Imposing a license requirement on all copyrighted works to protect a small fraction of rights holders would impose unsustainable transaction costs on researchers.
3. Finally, for works freely and lawfully available, there is very rarely any clear identification of what is protected and who owns it. So, imposing a licensing requirement on TDM research and machine learning would require negotiation of hundreds of thousands of potential licenses from unidentified owners around unclear rights. This would grind research to a halt and create the possibility of abusive copyright litigation against those engaged in research.

As such, **we would recommend an additional question** to be considered by WIPO's Secretariat:

- *Q: How can IP policy ensure licenses do not block access to and sharing of lawfully accessible data that could prevent the development of AI?*

### **Question 13.(vi)**

**Existing legislations have introduced safeguards to prevent any abuse** and potential copyright infringements related to text and data mining and machine learning.

For instance, as previously noted, the recently adopted EU Copyright Directive states that for content that is freely and lawfully available, copyright owners can use technical protection measures to prevent content from being crawled. However, it is **important that such measures be implemented only through machine-readable standards** (such as 'robot txt' format) and that strict and clear limitations to the reservation of rights by content owners be put in place, so as not to preclude the rights of the beneficiaries of potential copyright exceptions. For content that is not publicly available, rightsholders always have the possibility to use paywalls and licenses to reserve their rights.

WIPO has the opportunity to offer clarity to all TDM and machine learning users, **by adding the following question to its issues paper**:

- *Q: Should international IP policy encourage the use of machine-readable standards as the only acceptable standards for content owners to reserve their rights and prevent their content from being mined for artificial intelligence purposes?*

### **b) Copyright/Issue 9) General Policy Issue**

#### **Question 16**

As mentioned earlier in this contribution, **access to data and data sharing are crucial** to ensure artificial intelligence can flourish globally. Looking beyond exceptions to copyright, WIPO should also look at how international IP law could encourage access to data and data sharing by commercial, non-commercial and government entities, for commercial and non-commercial purposes.

EARE members **would recommend WIPO's Secretariat to include the following question** in its revised Issues Paper:

- *Q: In relation to Artificial Intelligence, how could IP policy be updated to support data sharing and pooling for the purposes of machine learning?*

### c) Data 10) Further Rights in Relation to Data

#### Question 23.(i)

**Current mechanisms in place today provide adequate protection for data.** The world is not data poor, there is no need to incentivize data generators to create more data, and the technological capacity to compile and distribute data are inexpensive and accessible. **Creating additional rights in order to protect data** beyond what is already available today does not seem warranted and **could actually impede innovation by placing unnecessary barriers to the use of data for developing artificial intelligence** and related technologies. A typical example is the EU Database directive of 1996 – in a 2018 [evaluation](#) of the directive, the European Commission found that while the directive’s limited scope facilitates its implementation, **the *sui generis* right should not be extended broadly to the data economy.** Indeed, the ability to deploy AI applications typically relies on the availability of wide and diverse data sets, which can be achieved through data sharing. Policy makers should consider the value of the outcomes that are achieved by using data and consider policies which encourage and enable the use of data – such as the recently adopted exceptions to copyright in in EU and other jurisdictions that have facilitated innovation in the area of AI, while protecting the commercial interests of content owners.

As such, we would recommend **WIPO’s Secretariat to include the following question** in its revised Issues Paper:

- *Q: Should new IP policies that facilitate data sharing and limit, not extend, the controls by which data generators can prevent the use of data, be introduced?*