

European Alliance for Research Excellence

Contribution to the UK's Intellectual Property Office's Consultation on Artificial Intelligence and Intellectual Property

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 [Allied for Startups](#), [BSA | The Software Alliance](#), [Research Libraries UK](#), [SCONUL](#) (Society of College, National and University Libraries), [UCL Library](#) (University College London) and [LACA](#) (UK Libraries and Archives Copyright Alliance), 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 contribute to the UK IPO's call for views on Artificial Intelligence (AI) and Intellectual Property (IP). Considering our remit spans TDM and Open Data, we will be focusing our contribution on questions 1 to 4 of the copyright and related rights section, and on questions 2 and 3 of the trade secrets section.

Copyright and related rights

1. *Do you agree with the above description of how AI may use copyright works and databases, when infringement takes place and which exceptions apply? Are there other technical and legal aspects that need to be considered?*

Text and Data Mining is a building block for both machine and deep learning, which play a foundational role in the development of AI. Collecting and analysing data using TDM is essential to enable machines to learn and AI algorithms to better recognise, understand, adjust, and respond. Without the appropriate legal and technical framework to permit a program, software or computers to access and analyse very large amounts of data and employ cognitive technologies to allow the learning of patterns, amassing sufficient amounts of data to develop, train and test AI would be challenging, and possibly prohibitive, for many individuals, researchers, and organisations.

TDM works by crawling thousands of different digital sources, and some of these can indeed be protected by copyright. However, as long as the TDM user has lawful access to a copyright-protected work they should be able to mine that content for purposes of machine learning. 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, and to communicate those findings, insights and observations to the public.

TDM and machine learning may also require automated, incidental storage of lawfully accessed copyrighted works to access non-copyrightable 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, and therefore do not constitute copyright infringement.

As rightly identified in the [UK Government's 2017 independent AI review](#), copyright exceptions for the purposes of Text and Data Mining are required to support the development and training of AI applications which offer vast potential for economic growth and increased competitiveness globally. We believe the current UK exception, covering only research purposes is too limited, and does not reflect the reality of today's research, which is often conducted as part of public-private partnerships. As it stands, TDM users falling under that category face a legal grey-zone, which tends to deter them from conducting TDM activities, for fear of copyright breach.

The UK could consider a number of legislative actions to encourage data innovation, while balancing economic expectations of rightsholders. EARE members have been advocating for a **broad, solid and easy-to-understand TDM exception in Europe**. In line with this position, we believe the **UK should broaden the current TDM exception** to cover private and commercial use. More specifically, the Copyright, Design and Patent Act could be updated to **allow for the reproduction of lawfully accessed works to facilitate TDM, for commercial or non-commercial purposes, by commercial and non-commercial entities**. In addition, a broadening of UK fair dealing exceptions to encompass a broad fair use concept, similar to that in the US would promote innovation and safeguard adverse economic impact to copyright owners.

2. Is there a need for greater clarity about who is liable when an AI infringes copyright?

EARE members are concerned that the formulation of this question assumes that the use of data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright. 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.

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, used to train AI, should not be covered by copyright.

As such, we believe a discussion on infringement and liability is premature until there is further clarity on what constitutes a copyright infringement by an AI.

3. Is there a need to clarify existing exceptions, to create new ones, or to promote licensing, in order to support the use of copyright works by AI systems? Please provide any evidence to justify this.

As mentioned earlier, the right exceptions to copyright to conduct Text and Data Mining are required to support the development and training of Artificial Intelligence applications. EARE members believe UK 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 lawfully accessed or acquired data** – such as the recently adopted exceptions to copyright in the EU and other jurisdictions that have facilitated innovation in the area of AI, while protecting the commercial interests of content owners.

We believe it is critical that the UK Government encourages and fosters an environment where TDM is not unnecessarily burdened and remains accessible to all entities that already have lawful access to data they wish to mine, for all purposes. 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.**

For instance, Articles 3 and 4 of the EU Copyright Directive provide the appropriate framework to encourage the development of AI in Europe and bring much needed clarity for all entities that are at the forefront of the research and innovation ecosystem in Europe, including researchers, libraries, startups, SMEs, technology companies, and others.

On the other hand, 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 vast quantities of data. Text and data mining, which works by crawling thousands of different digital sources, is a way to obtain large amounts of public data for the purposes of training models and AI. As long as the TDM user has legal access to a copyright-protected work, either through a license or because they are publicly accessible, 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.
 4. Making available open data sets, in a manner that is accessible and usable by the public, while safeguarding privacy and other human rights, can also increase the quantity and quality of data available to researchers and organizations for the purpose of training machine learning models and algorithms. This may include making available public sector data, consistent with data sharing norms and laws and for the private sector to contribute data sets that do not include private, sensitive, or confidential information.
4. *Is there a need to provide additional protection for copyright or database owners whose works are used by AI systems? Please provide any evidence to justify this.*

We strongly believe that existing legislations have introduced sufficient 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. 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 technical measures, such as paywalls and authentication mechanisms to reserve their rights.

Trade secrets

2. *Does the nature of AI pose any problems if UK trade secret protection is required? Does UK trade secret law give adequate protection to aspects of AI technology where no other intellectual property rights are available?*
3. *What are the advantages and disadvantages of using trade secrets in the AI sector? Could information that is not shared inhibit AI development?*

As a coalition, EARE members believe that **trade secrets laws, as they exist today do not create a significant barrier to AI development.** EARE strongly supports the principle that data should be "as open as possible, as closed as necessary", and that when it comes to data, sharing is winning. Companies should be able to protect the data that they consider to be commercially-sensitive should they wish to, using current laws on trade secrets, but should also be encouraged to share them in a safe and trusted way if they wish to, through contractual agreements for instance.

We believe that enough tools exist today to promote the sharing of privately held data, sometimes protected by trade secrets, whilst retaining return on investment, such as innovative governance structures (data trusts, data cooperatives or data commons), effective legal agreements that facilitate access to data (data sharing agreements), and the use of technologies that help protect privacy and strengthen security.