

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](#), LIBER Europe, [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 consultation on Artificial Intelligence (AI) and Intellectual Property (IP): copyright and patents. Considering our remit spans TDM and Open Data, we will be focusing our contribution on the second area explored by the consultation on "*Licensing or exception to copyright for text and data mining, which is often significant in AI use and development*".

Overleaf is the response form filled in.

Annex - Response form

Section A

Copyright – text and data mining (TDM)

1. *If you license works for TDM, or purchase such licences, can you provide information on the costs and benefits of these? For example, availability, price-point, whether additional services are included or available, number and types of works covered by the licence etc.*

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

1. Machine learning relies on 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 and startups**, and therefore hinder innovation in the UK. For instance, some of our members had to rethink the way they were conducting research projects because they were unable to afford licenses to access data in a usable form (via an API provided by the platform), as these can amount to up to more than USD 30,000. In order to circumvent this barrier, it often happens that researchers end up downloading materials they have legal access to in bulk as PDF documents and then process these documents outside of the platform. This considerably slows down research projects and subsequent innovations.
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.

2. *Is there a specific approach the government should adopt in relation to licensing?*

As outlined in question 1, we believe that licensing for the purposes of text and data mining and machine learning are very problematic. Licenses for these purposes will only deter innovation and encourage innovators to move to more innovation-friendly geographies. Therefore, to promote AI innovation, we recommend the UK Government ensure that any updates to its IP policy clearly state that the lawful access to and use of data for computational analysis is not subject to additional licenses. use of licenses should not block access to and sharing of data and prevent that could prevent the development of AI.

3. *Please rank the options in order of preference (most to least preferred) and explain why.*

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 and **opt for option 4**. 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.

We believe that the vast majority of rights-holders will not object to the use of their works permitted by a broader exception. More importantly, any update to the UK Copyright law should recognise that there are multiple ways rights-holders and users can derive benefit from copyrighted material that do not involve traditional restrictive licensing for payment, such as the use of permissive open source licenses in the software development area. We would urge the UK Government updates its Copyright law in a way that encourages a diverse array of permissive use relationships that serve the needs of the broadest set of rights owners and users.

TDM exceptions should reflect the realities of 21st century research, where the growing use of big data and artificial intelligence tools in research and innovation now result from the seamless collaboration between public and private organisations. For all these reasons, **EARE members favour option 4**. Other options are ranked below.

Text and Data Mining (TDM)	
Option 4	Adopt a TDM exception for any use, which does not allow rights holders to opt out
Option 3	Adopt a TDM exception for any use, with a rights holder opt-out
Option 2	Extend the existing TDM exception to cover commercial research and databases
Option 1	Improve licensing environment for the purposes of TDM
Option 0	Make no legal change

4. *If you have experience of the EU exception with opt out for rights holders, how has this affected you?*

Based on our members experience with the EU exception, the **ability of rights holder to “opt out”** under Article 4 of the relevant EU Digital Single Market Directive is **making the EU less competitive** in comparison to the US, Japan and other jurisdictions where no such opt out for commercial use exists.

We believe that strict and clear limitations to the reservation of rights by content owners should be put in place. Maintaining a sufficient level of protection for content owners should not preclude the right of the beneficiaries of these exceptions, both commercial and non-commercial entities, to conduct data mining activities on lawfully accessible content without the risk of being locked-out by the disproportionate use of technical protection measures or the need to interpret ambiguous contract terms. Moreover, often, TDM is carried out in an automated manner via bots, that cannot read plain language terms and conditions. Therefore, we **would strongly welcome language requiring that**

content owners implement existing machine-readable standards recognised by standard online technologies (such as robot.txt) to reserve their rights and prevent TDM activities on lawfully accessible works, in addition to explicit language in their online agreements.

Allowing opt-out through other means would create considerable hurdles for smaller actors of the innovation ecosystems, who will remain in the current legal grey-zone, which tends to deter them from conducting TDM activities, for fear of copyright breach. 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.

5. How would any of the exception options positively or negatively affect you? Please quantify this if possible.

Other countries in the world have made it legally easy to conduct TDM to help spur the pace of innovation and research, leading some European innovators to work abroad. **United States** courts, for example, have repeatedly recognised that non-expressive use of materials such as TDM help promote copyright's goals of expanding access to information and learning without harming legitimate copyright owner interests. **Japan** also [updated](#) its "Copyright Act" in 2018 to promote innovative digital and AI services. With this update, Japan successfully managed to balance their copyright rules to support their technological ambitions, thereby allowing researchers and private companies to carry out machine learning activities, primarily by removing ambiguity for using copyrighted works for understanding and analysis. Similarly, **Singapore** amended its copyright laws to allow text and data mining by all entities to help promote Singapore's ability to participate and compete in the emerging data analytics industry. **China** is also aiming to become the world's primary AI innovation centre by 2030, and has recently supported updating their copyright laws to include an express TDM exception.

All these countries are well on track to achieving their common ambitions to become "AI-powerhouses" by massively investing in the development of these technologies and by offering a welcoming legal environment to data analytics firms, researchers and innovators.

On the other hand, 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 (option 0), 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.** The [testimony](#) of [REDACTED] clearly shows the limits of such an exception to boost innovation. Similarly, adopting an exception as outlined under options 1 and 2 would remain too limited and would not put the UK on par with the legislatures that have adopted broad exceptions, like options 3 and 4.

Deciding to update UK TDM rules by opting for options 0, 1 or 2 will deter innovators from conducting TDM-related activities in the UK, as many other geographies have adopted AI-friendly copyright rules. From our experience, **the only option under which the UK Government would realise its ambitions to become a global "AI-powerhouse" is option 4.**

--

Section B: Respondent information

A: Please give your name (name of individual, business or organisation).

European Alliance for Research Excellence (EARE)

B: Are you responding as an individual, business or on behalf of an organisation?

- 1) Organisation – please provide the name of the organisation
European Alliance for Research Excellence

C: If you are responding on behalf of an organisation, please give a summary of who you represent.

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](#), LIBER Europe, [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.

E: If you are responding on behalf of an organisation, are you?

- 1) Any other type of organisation - please specify

EARE is a coalition of companies and research organisations that are committed to the future of innovation and R&D in Europe, including in the UK.

F: If you are responding on behalf of a business or organisation, in which sector(s) do you operate? (choose all that apply)

- 1) Information and communication – Telecommunication
- 2) Information and communication – IT and another Information Services
- 3) Scientific and technical activities
- 4) Education
- 5) Other activities – please specify: Research and innovation organisations

G: How many people work for your business or organisation across the UK as a whole?
Please estimate if you are unsure.

- 1) Fewer than 10 people

H: The Intellectual Property Office may wish to contact you to discuss your response. Would you be happy to be contacted to discuss your response?

Yes.

I: If you are happy to be contacted by the Intellectual Property Office, please provide a contact email address.

[REDACTED]

J: Would you like an acknowledgement of receipt of your response? Yes/No

Yes.