EARE contribution to the UK Consultation on AI and Copyright

As part of the consultation launched by the UK Government on copyright and AI, EARE supports an **approach that promotes innovation and research excellence**. It is crucial for the UK Government to seize its unique opportunity for global leadership in digital innovation. AI drives this innovation in all sectors of the economy, and **intellectual property should foster innovation instead of limiting it. Copyright should not be used to solve issues raised by the development of AI.** Drawing on our experience within the EU, we have developed several recommendations for the UK government.

Our vision for the UK's future: Strengthening AI leadership through open data policies and research excellence

- The UK's AI Opportunities Action Plan sets out a clear strategy to ensure British leadership in AI development. To achieve this goal, the UK must establish a legal framework which **fosters innovation and research excellence**. This is the only way to remain competitive with AI leaders such as the US and China. In developing this policy framework, the UK should take inspiration from countries, like Japan and Singapore, which have adopted **pro-innovation copyright laws** that strongly **support machine reading** and **minimise bias**.
- The UK currently invests **over £17 billion a year in research and innovation**, to turn the latest science into new cancer treatments and reduce energy consumption to tackle climate change. These important societal transformations will only be possible if the UK government implements **policies that ensure research excellence and innovation**, **facilitate access to data**, and **develop advanced AI systems**.
- EARE welcomes clarification of the current UK's copyright law to provide **broad data mining exceptions (Option 2)**, which would enable fair access to data for all. This option would **reduce research bias**, **drive innovation**, and **simplify the current complex legal landscape**.
- In contrast, Option 3, the Government's preferred approach, **would introduce a relicensing scheme.** This would **increase the burden on all British businesses, hitting researchers and innovators the hardest**. This system would prevent them from using content they already have access to, unless they make a new agreement for AI purposes. It would also require them to renegotiate access to content for AI purposes that they already have access to, reducing the volume of data available for AI models, increasing bias and imposing costs that researchers and innovators in the US and East Asia do not face, making the UK less attractive as a place for research and innovation. It is important to note that countries such as the US, Japan, or South Korea, which follow a more science and innovation-friendly approach to copyright, continue to have a thriving creative industry. Instead of encouraging relicensing through a system that takes inspiration from the EU's approach, **the UK government should encourage flexible exceptions and open data policies to ensure a competitive AI landscape** and prevent further excessive market consolidation. Similarly, technical solutions for implementing the reservation of rights in the EU, which may be used to accommodate option

3, remain complex and, at present, are neither standardised, nor interoperable. The future changes in the legislation should remain technically feasible and economically viable, particularly for start-ups, researchers, and innovators. **The modernisation of copyright law in the UK should not be delayed by the development of technical mechanisms for respecting the opt out**, which can progress independently. The EU currently enjoys a TDM exception, and the UK cannot afford to be left behind.

- **The UK must also recognise the important role of researchers and innovators** in the economy and the impact relicensing would have on them. **Changes to copyright law would affect all sectors reliant on data** for their day-to-day work, including the medical, agricultural, and STEM industries. Any restrictions to data access would have far-reaching economic and scientific consequences. Public fund research will be prevented from being fully leveraged for the public good, and licences will delay the potential impact of research activity, adding significant costs, and lowering the return on investment by the UK government.
- **The UK government should also consider the importance of education and the need to develop a highly digitally literate workforce**. Training and educational institutions need to be able to develop, build, deploy, and release AI related tools to make it possible.
- To ensure a balanced system that supports both rights holders and the creative industry, the UK Government should consider targeted policy instruments that address the needs of creators without harming research and innovation. **Copyright should not be used to solve issues raised by the development of AI**. As a comparatively small economy with no first-mover advantage, the only way to make economic and societal progress together is through open, accessible, transparent, reproducible, and cooperative research and innovation.

OUR APPROACH

Our preferred option would be Option 2: 'A broad data mining exception'

- **EARE members welcome clarification of the UK's current copyright law**. Existing exceptions in UK copyright law, such as Section 29a and the temporary copyright exception, permit AI developers and researchers to use copyrighted works for training purposes. However, the current legal framework remains uncertain. While non-commercial researchers benefit from a broad Text and Data Mining (TDM) exception, they often operate within public-private partnerships (PPPs), receive mixed funding, or have goals that are not strictly non-commercial. This makes their projects more complicated, as purpose-based restrictions on data mining limit their ability to make the most of this exception through innovation and broader societal benefits. Additionally, due to licensing requirements for commercial research, both non-commercial and commercial researchers and innovators often face hurdles to accessing content they already have legal access to.

- As part of this reform, we believe that data should be "***as open as possible, as closed as necessary***" for the UK to leverage the full potential of its research and innovation capacities.

- Therefore, EARE favours **a broad data mining exception** (*Option 2*) as the **best way to ensure fair access to data for all**, enabling the development of competitive AI models in the UK. This broad data mining exception would **minimise biases in training AI**, ensuring research excellence in the UK, as data will be accessible for any purposes, including commercial purposes.

- **Requiring legal access to data used to mine algorithms would ensure respect for rights holders' interests,** in line with the principle "***the right to read is the right to mine***". Permitting organisations the ability to train on and analyse content they have legal access to will not only **minimise bias** arising from not having access to enough of the right kind of data (volume and veracity), it will also mean that deadweight **costs of relicensing that many UK competitors do not have** will be avoided. Such an exception would also **prevent the creation of gatekeepers** or market consolidation which would significantly stifle innovation and create imbalances.

- **This new approach would put an end to the artificial separation of commercial and non-commercial purposes.** This distinction is increasingly difficult to justify in modern research contexts (PPPs, mixed funding, job sharing, etc) but also causes uncertainty when a research project leads to commercial products, as a result of knowledge transfer. Several academic projects are currently using AI to accelerate research and benefit UK society e.g. Exscientia, which originated at the University of Dundee, and is using AI to accelerate discovery of new medicines. Exscientia's approach has led to partnerships with major pharmaceutical companies. Consequently, distinguishing between the purposes **limits the ability of researchers to apply findings** in real-world, impactful ways. The UK should learn from the EU approach's weaknesses, where the distinction between commercial and non-commercial purposes led to many uncertainties about how it works in practice. Moreover, there is a growing concern around the level of opt-outs and how this contributes to bias and comparatively poor model prediction in domain specific areas of AI development. Additionally, the possibility of opt-out significantly weakens the universality of Article 4 of the Directive on Copyright in the Digital Single Market. **EARE urges the UK to avoid replicating a system that has shown significant weaknesses in practice.**

- **Copyright rules governing AI-generated content would affect all sectors of a research-intensive economy** as AI is increasingly used in areas ranging from healthcare, finance, education and entertainment, and in particular STEM industries while primarily benefiting a small segment of the society, namely the creative industry. If introduced, the right to opt-out would significantly undermine AI development creating inefficiencies that result in valuable data being left unused, while also creating barriers to fair competition. When rights holders opt-out of TDM exceptions, their data cannot be used for AI and research, unless other exceptions apply. Such data goes unused and does not generate revenue for the rights holder. In the event this data is useful for

certain applications, a licence would have to be negotiated – entailing high fees and complex negotiations that can be prohibitive for startups, SMEs, and researchers. This puts them at a competitive disadvantage against large companies with substantial resources and leads to market consolidation. It also contradicts the UK Government's rationale for research funding through, amongst others, the research councils – namely, the commercial application of basic research.

- **It is imperative that AI tools are accurate, reliable, reproducible, and free from bias – with data as the backbone that underpins this**. Increased access to broad and varied data should therefore be prioritized to avoid bias, poor model prediction and unlock AI development and benefit researchers and innovators. Policy interventions such as the Option 3 preferred by the UK government can shape socioeconomic biases on AI development. As mentioned in the latest publication from the US Copyright Office, developers may rely on public domain works to mitigate the risk of copyright infringement.[1] As a result, the training material may reflect privileged historical socioeconomic norms rather than contemporary values, introducing biases into AI models. Similarly, when using copyrighted works, AI developers may favour large datasets controlled by a few rights holders to reduce costs. This can exclude diverse and valuable data, resulting in unrepresentative training material and reinforcing biases in AI outputs.[2] To avoid bias and develop accurate AI models, the UK should promote open data policies, data access, and infrastructures that facilitate data sharing while maintaining security. In this context, the UK should also explore Secondary Publication Rights which would enable immediate access to a wide range of publicly funded research covering all types of uses.

- Countries such as Japan and Singapore have passed pro-innovation copyright laws that support machine learning and minimise bias, while US case law considers that the analysis of data constitutes fair use. Japan allows TDM for any purpose (as do the USA and most countries) with no distinction between commercial and non-commercial purposes, as long as it does not conflict with the normal exploitation of the work or unreasonably prejudice rights holders. Therefore, only a flexible exception (option 2) would support the Prime Minister's efforts to position the UK as an AI "superpower". It is also important to note that countries such as the US, Japan, or South Korea, which follow a more innovative approach to copyright, continue to have a thriving creative industry.

DEEP-DIVES INTO FOCUS AREAS

Relicensing and opt-outs should be avoided, as this stifles innovation and competition

- **As explained above, relicensing practices which follow opt-outs would severely hinder competition**. Such practices disadvantage smaller AI players and research organisations with a lack of resources. Indeed, when individual users or organisations opt-out and renegotiate access to

[1] Lutes, Brent A. ed., Identifying the Economic Implications of Artificial Intelligence for Copyright Policy: Context and Direction for Economic Research, U.S. Copyright Office, 2025.
[2] Idem.

material an organisation already has legal access to, this creates barriers to competition and a dynamic where large companies have the resources to negotiate and can maintain their dominance, leading to further market consolidation. However, for innovators and researchers, obtaining licences for machine learning can represent a significant challenge. According to our members, at the start of a project, many of the questions necessary to secure such licences cannot be fully answered.[3] As a result, researchers and startups often struggle to obtain the required permissions. This situation contrasts with the framework in countries such as the US, Japan, or Singapore where this negotiation process is not required. Instead, promoting broad data mining exceptions can foster healthy competition and innovation in the AI landscape by keeping valuable datasets accessible to a wider range of developers, researchers and smaller AI companies.

- **The legal access option effectively balances the interests of researchers, businesses and rights holders as it prevents unrestricted web scraping or data mining on copyrighted materials without authorisation.** To perform TDM, users will be required to have legal access to the content, such as through open access data, the internet, subscriptions, purchases, or other forms of legitimate access. This means researchers, businesses, and other TDM users already pay for access to the content, either directly (e.g., subscription fees) or indirectly (e.g., library funding for journals). These subscriptions already represent a substantial investment, and smaller entities often cannot afford both the costs of access and additional licensing fees. In addition, we estimate that the UK taxpayer already pays in the region of £20 billion per year to fund research and then buy it back in the form of journals, books, etc. This **publicly funded research** should be legally accessible to all, given that it is taxpayer funded in the first place. We are also deeply concerned that a requirement to relicense content an organisation already has legal access to, will make scientific publishers' gatekeepers of downstream data-driven scientific innovation. They will be able to decide which AI-based research and models can be developed and which not. For example, they will be able to decide which diseases, medical applications, and environmental technologies can be researched using AI technologies and which cannot. This is not what copyright was intended for.

- **Consequently, we call on the UK Government to support the principle that the right to read is the right to mine**. Analysing data through TDM is a transformative use that should fall under the original access rights rather than requiring a separate license. Therefore, additional **relicensing** mechanisms should not be required for mining data that is already accessible. EARE strongly supports open access policies and open science and opposes additional hurdles to secure mining rights for data one already has access to.

[3] Knowledge Rights 21. Example Case Study of Licence Negotiations for Machine Learning.

Collective licensing

- **Government involvement in encouraging collective licensing will be counterproductive and should be avoided.** This question is predicated on the assumption that Collective Management Organisations (CMOs) have digital rights to allow data analysis. As far as we are aware, no CMO currently has such rights given that most rights holders do not grant UK CMOs many digital rights at all. Furthermore, such involvement would introduce additional layers of negotiations and compliance, delaying access to datasets for training AI. Mandatory licensing fees and royalties could make acquiring training data more expensive, which might hinder the development of AI models by developers, researchers, and innovators. Furthermore, when publishers opt-out as mentioned above in important areas such as science it will potentially cause bias and weakened models which must be avoided at all costs.

- **In the context of opt-outs, even if CMOs were mandated by the UK Government or granted such rights we consider that they should not be entitled to exercise a reservation of rights, notably through the Extended Collective Licensing (ECL) framework.** This would significantly encourage the adoption of opt-outs and would essentially result in an opt-in system in the UK, with a systematic requirement to obtain another licence to engage in any TDM activity for works managed by a CMO, as well as the entire repertoire of works. This would prohibit AI development in the UK and significantly disadvantage UK AI developers and researchers.

- **Instead of encouraging licensing, the UK Government should focus on data access and promote open data policies to drive innovation.** If the UK chooses Option 3, it should encourage data sharing instead of encouraging opt-outs and collective licensing to ensure AI development. This approach would reduce barriers, democratise information, enable broader participation in innovation and research, and foster economic growth and scientific discovery.

Use of AI in education

- Rather than focusing on creative professionals, the creative industries and AI developers, the UK Government should take a wider perspective. AI tools and research are and will be key to understanding and addressing societal challenges, such as climate change or pandemics. Instead of limiting AI development to meet the demands of a small segment of society, **we must support cooperation between institutions as well as knowledge transfer, encourage interdisciplinary knowledge-sharing, create social and economic value by transforming data and research results into solutions, and foster AI uptake in science**.

- **A multi-faceted strategy is required to make this a reality, including policy initiatives, funding, incentives to ensure data access, and education.** As part of this strategy, the UK should encourage voluntary data sharing by providing incentives for individuals and organisations to share data and educating them about the societal benefits of data sharing.

Transparency

- **Transparency should avoid additional administrative burdens and respect trade secrets**. Transparency is a core value and we welcome any transparency measures that aim to ensure trust and AI uptake. However, the suggested transparency measures can introduce additional compliance costs for research organisations and innovators, especially startups and SMEs. As recently mentioned in a European Commission study, obligations such as those established in the AI Act, requiring detailed summaries of the data used for training, can "*add a layer of compliance costs for research organisations*". They can also significantly affect trade secrets and intellectual property rights. Any **transparency measures should be voluntary, easy to implement, accessible, and avoid any excessive burdens** or excessive reporting for researchers and innovators. Complex documentation and the disclosure of sensitive data should be avoided.
- **In the EU, the current implementation of the transparency requirements included in the AI Act face several challenges** and are often viewed unpractical or unworkable, especially for smaller players. The UK Government should learn from the EU's approach to transparency, avoiding the creation of an overly burdensome system.
- **The current direction of transparency requirements included in the AI Act will undermine knowledge transfer activities and chill non-commercial research undertaken under Art 3 of the Directive.** Transparency requirements on commercial entities, unless proportionately implemented, will undermine public private partnerships. There is a danger that scientific publishers will use the obligations in the AI Act on commercial firms who have worked with public organisations to force retrospective transparency regarding models developed under Article 3 that are not subject to the AI Act. This is indicative of the need to think carefully how transparency obligations will affect all actors, including universities and charities.

Encouraging research and innovation

- **As explained above, the existing TDM exception for non-commercial research is not fit for purpose. We call on the UK Government to remove the distinction between commercial and non-commercial purposes.** This distinction is increasingly difficult to establish in the modern research context (PPPs, mixed funding) but also leads to uncertainties when a research project leads to commercial products. Consequently, distinguishing between these purposes limits the ability of researchers to apply findings in real-world, impactful ways as required often by funders such as UKRI, ARIA etc, such as developing AI tools, improving healthcare solutions, or addressing other societal change.

- **Section 29A and the introduction of a flexible exception like in Japan or fair use should be accompanied by an explicit recognition of the fact that digitisation and the sharing of data amongst project partners is permitted.** Furthermore, the UK should amend its pre-Brexit circumvention of Technological Protection Measures provision and allow organisations

undertaking research to circumvent TPMs without the need for permission from the UK government.[4]

- **No distinction related to purposes, or the size of an AI firm, should be considered**. The purpose of an AI system is not always clear-cut and often difficult to legally define. Additionally, a system where size does not matter encourages collaboration across industries and sectors. Smaller firms and researchers can work alongside large companies and universities, leading to synergies and more diverse innovation. Enforcing distinction requires extensive and complex monitoring and compliance systems, often leading to administrative burdens for researchers and innovators, additional costs, and barriers to innovation.

AI outcomes: Protection for the outputs of generative AI

- **Protection for computer-generated works is unnecessary**. EARE supports open access to data and therefore, considers that such protection could become a barrier for researchers, innovators, and developers to access data. Open data policies, which promote the availability, accessibility, and usability of data should be promoted for a more innovative AI landscape in the UK, and this includes AI generated works such as synthetic data.

EARE'S KEY RECOMMENDATIONS

In summary, EARE proposes the following recommendations to ensure a balanced approach to UK's copyright and AI policies:

1. **Clarifying UK copyright law for broad data mining exceptions (Option 2)**
    - We support **clarifying the UK's copyright law to ensure broad data mining exceptions (Option 2)**, enabling fair access to data for all stakeholders. This will help reduce bias, foster innovation, and address the complexities of the current legal framework.

2. **Prioritize data sharing and knowledge valorisation if option 3 is preferred**
    - If option 3 is preferred, the UK government should **prioritise data sharing and knowledge valorisation** to make sure that the current investment in research can benefit UK society.
    - This new approach should **remove the distinction between commercial and non-commercial purposes and tackle the current conflict between copyright policy and the UK's innovation policy which invests over £17 billion a year in science.**
    - **A reservation of rights should waive the availability of the TDM exception, and should not override other exceptions that may apply**

[4] For further information on the problems created by TPMS for researchers see the following three studies on TPMs: Kristofer Erickson and Victoria Stobo, Survey on Technological Protection Measures: Impacts for Researchers, Libraries and Archives (2024), Project Report, Knowledge Rights 21 (doi: 10.5281/zenodo.14168677). Kristofer Erickson and Felix Rodriguez Perez, Technological Protection Measures and Digital Preservation. Evidence from Video Games; Anthony D. Rosborough, Technological Protection Measures and the Law, Impacts on Research Education & Preservation.

- The UK Government should also **reject the idea that CMOs may exercise a reservation of rights** or conduct licensing for TDM activity on behalf of members or non-members.
- Instead of encouraging relicensing, the UK government should encourage **flexible exceptions** and **open data policies** to ensure a competitive AI landscape and avoid market consolidation.
- The UK government could encourage content creators to allow their works to be used in AI training. For example, offering access to AI tools for their own works. Moreover, the UK government can also develop guidelines and best practices or implement educational campaigns on AI training.

3. **Ensuring balanced transparency obligations**
    - Transparency requirements must be **proportionate, respecting trade secrets**, while supporting non-commercial research and knowledge transfer.
    - Transparency obligations should be **practical and with limited overheads** to ensure AI development in the UK and avoid giving scientific publishers the ability to undertake disruptive "fishing trips" in the upstream non-commercial activities of British universities.

About EARE: *The European Alliance for Research Excellence (EARE) was convened in 2017, and now brings together eight members from the research and innovation ecosystem in Europe, including BSA | The Software Alliance, Microsoft, Allied for Startups, LIBER, LACA, Research Libraries UK, SCONUL (Society of College, National and University Libraries), and UCL (University College London) Library, advocating for the EU to live up to its innovation potential in the digital economy.*