Libraries and Archives Copyright Alliance (LACA)

Response to UK Government Consultation on Copyright and AI, February 2025

Overview of LACA's position

The Libraries and Archives Copyright Alliance (LACA) comprises a range of member organisations including university and museum libraries and archives, NHS education bodies, private charitable organisations, groups acting on behalf of libraries and archives such as Research Libraries UK (RLUK), Society of College National and University Libraries (SCONUL), the UK's National Libraries, groups representing the needs of disabled users, UK university libraries and more.

We work on behalf of our sector and our users, in particular those requiring access to digital services such as digitised collections, research data sets, digital libraries and eBook resources.

The topic of this consultation is a contentious one, and our members are working hard to understand the implications of generative AI for our organisations and our users. We recognise how crucial regulation is to this emergent technology to create an environment in which creativity and innovation can thrive while also leveraging the transformative potential of AI for the good of society.

LACA's response is based on the view that research is the engine for social progress, and any regulation needs to factor in the multitude of ways that AI impacts the research sector.

To this end LACA is concerned on a number of levels by the implications of this consultation. When developing the copyright rules that govern data-driven innovation the government needs to consider the following:

- If detrimental harms affecting commercial creators and intermediaries are proven, **targeted cultural policies should be adopted**, not laws like copyright which affect innovation and scientific innovation more broadly, potentially negatively impacting on the goals of our sector.
- AI investment will take place where laws governing the reuse of data / information are most flexible. Option 3 sets out a relicensing regime, requiring organisations to relicense access to content that they already have legal access to. **This deadweight cost to UK organisations will make UK-developed AI uncompetitive and limit the benefits to our sector.** In this respect we refer you to the submission by the UK Day One Project which estimates Option 3 will cause a **loss** of at least £17.67bn to the UK's GDP over the next 10 years.
- The UK government's investment of tens of billions a year in publicly funded scientific research and its commercial application by research intensive industries will indelibly be undermined by Option 3.
- Overly burdensome regulation applied to lawfully-accessed data for training AI models will reduce the content a model can train on (both volume and therefore veracity) compared to models developed in flexible copyright regimes. This will inevitably mean **models developed here are less able to make accurate predictions, and more likely to contain bias**, negatively impacting our sector.[1]
- The consultation appears to stretch what copyright regulates to cover issues as wide-ranging as personality rights, transparency, and automated copying by computers for non-communicative purposes - something which falls outside of the scope of copyright protection. We believe, as outlined herein, that **weakening an already over-stretched copyright regime will have detrimental effects not only on the law, but on the research, educational and information sector that benefits society and economy as a whole**.

B.4 Policy Options

Question 1 Do you agree that option 3 is most likely to meet the objectives set out above

Option 3 is unlikely to meet the stated objectives.

By framing the recommendations as a path to resolve a false dichotomy between the interests of the creative industries and tech companies, the government misses the educational, research and scientific innovation sectors (e.g. STEM, medicine, NHS etc) that both underpin development of AI technologies, and provide social, cultural and economic value by building on or incorporating AI models and applications.

[1] Amanda Levendowski, 'How copyright law can fix artificial intelligence's implicit bias problem' (2018) Washington Law Review 93

We recognise the government's role in implementing certain controls and transparency measures to ensure the safety of users, to protect certain copyrighted works from commercial exploitation, or to provide guardrails around the social and economic impacts of AI technologies. But we believe these measures should be proportionate to the risks. Control mechanisms need to be flexible, efficient and forward-thinking enough to keep up with the pace of change we are currently experiencing in the realm of AI development.

China's DeepSeek is a recent obvious example of the speed of development. DeepSeek is based on an AI model comparable in quality to OpenAI and Google's DeepMind, but developed at a fraction of the cost.[2] This company is less than 2 years' old and is already shaking up the tech world. The open nature of DeepSeek provides an important economic lesson for UK policy makers. As a country, the UK does not have first-mover advantage in AI markets, neither does it have any large technology companies that are investing billions of dollars in AI models and their deployment across the globe. As a country it seems obvious that we can only compete with market incumbents by following strong open software and open data policies. This requires us to pursue flexible data regimes that maximise access to content for AI purposes, and therefore demands a flexible approach to copyright exceptions.

The Government's **AI Opportunities Action Plan**[3] aims to support "strong fundamental AI research, and high-quality research and engineering talent coming out of our universities", and to support the public sector as well as "encourage the private sector" to develop AI products and services. This Action Plan recognises the crucial role that the private sector plays in bringing about this vision of AI for the UK, working in collaboration with the public sector. This opportunity would be severely undermined should the copyright controls envisioned here in relation to the creative industries be applied as a "one-size-fits-all" approach to all aspects of the digital economy, research and innovation landscape.

The UK is strong in STEM research and data analysis, data production and information management - advantages our members strive to enable and foster. We believe the recommended approach could, if implemented as described in Option 3, prevent our member organisations and their users from reaping the full benefits of AI.

Question 2 Which option do you prefer and why?

Option 2: An open norm supplemented by a broad data mining exception.

The UK's reliance on a prescriptive list of copyright exemptions fails to support research, innovation, and scientific and technological advancement. The addition of a **flexible and open-ended exception**, as we see in technology-oriented countries in Asia and the United States, would allow the UK to advance creativity and research and benefit from technological and scientific innovations.[4]

---

[2] https://www.ft.com/content/d72e0750-6a8b-4ef4-b9e1-6d35fd2a69b8
[3] https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan
[4] Mendis, Dinusha and White, Benjamin and Hong, Dukki, Copyright and Open Norms in Seven Jurisdictions: Benefits, Challenges & Policy Recommendations (February 15, 2024).https://ssrn.com/abstract=4728782

At the same time, beneficiaries of exceptions need detail and clarity in order to make informed and confident decisions. To this end, we would encourage the UK government to **broaden the current exceptions for temporary copies (s28A) and for text and data mining (s29A CDPA)**. Our proposed approach on these can be found under Q26, 27 and 28.

C.1 Exception with rights reservation

Question 3. Do you support the introduction of an exception along the lines outlined above?

No. We think that the government should allow for the reproduction of lawfully accessed works to facilitate text and data mining (TDM) activity for commercial and non-commercial purposes by implementing a **flexible and open-ended exception**, backed up by the broadening of s28A and 29A CDPA as described in detail in Q26, 27 and 28.

Given that we represent a user base for library, archive and information services, we do not think that rights holder opt-outs would be practical or desirable for our constituents, and that such a regime would stifle research activity, knowledge transfer from universities, public-private partnerships and innovation more broadly.

Although the EU's Digital Single Market Directive Article 4(1) appears to permit reproduction and extraction for data mining, including if there is a commercial purpose, it allows right holders to opt out of the exemption. This puts for-profit TDM at the mercy of right holders, and introduces additional complexity and uncertainty. The transactional burden of getting rights clearance is heavy to the point of suppressing TDM activity. This burden is disproportionately heavy on start-ups, SMEs, public-private partnerships and researchers – particularly those not affiliated to an organisation.

In addition, the government should provide a clear mechanism for redress where technical protection measures inhibit such activities.

In LACA's view, opt-outs are not a desirable approach. The UK spends circa £20 billion each year in funding research for its downstream adoption by industry amongst other actors. Billions are then spent buying this research back. By introducing opt-outs and the relicensing of information that organisations already have legal access to, we will stymy innovation, preventing our sector from reaping the benefits of AI. These include:

1. The introduction of a relicensing regime that would compel organisations to obtain permissions to use content for which they already possess legal authorisation will add **significant overheads** for research organisations, including those involved in public-private partnerships as well as significantly **increasing the cost of licensing**. We believe such costs would be far beyond the means of universities, university spin-outs, start ups, SMEs etc.

2. The introduction of a relicensing framework, requiring organisations to relicense access to content to which they already hold legal rights, would inevitably **reduce the volume and diversity of data available for AI training**. The resultant limitations would not only impair the predictive accuracy of AI models developed within the UK

but also exacerbate biases within these models relative to those trained in jurisdictions with more flexible copyright regimes.[5]

3. Given the tens of billions of taxpayers money that are invested in research each year, we are deeply concerned that **data repositories and publishers will be able to control via Option 3 what commercial and knowledge transfer-related research can be undertaken in the UK**. It is difficult to conceive what the advantage to the UK would be of making such organisations gatekeepers of what research and commercial application of science can be made by commercial entities downstream. Taking an example, a publisher will be able to dictate whether certain diseases can be researched using machine learning technologies and whether some cannot according to the price they set, the terms and conditions they impose and whether exclusive licences have already been entered into. For a more detailed analysis of this we refer you to the **Knowledge Rights 21 submission**.

Question 5. What influence, positive or negative, would the introduction of an exception along these lines have on you or your organisation? Please provide quantitative information where possible.

Please see answers above. Furthermore, in terms of copyright jurisprudence we are concerned that **copyright is ill-suited to many of the suggestions outlined herein.** Copyright's sole goal is to protect against substantial similarity between works. We advise against stretching copyright to deal with issues for which it is ill-equipped for and not designed to address. Thus, any remedies required that fall beyond substantial similarity need to be brought to bear through targeted cultural policies – not copyright law - in order to mitigate the chilling effect overbearing regulation could have on our sector's goal to support and foster innovation and research.

We also would like to repeat that - as the **Authors Alliance submission** points out - the interests of public-minded research, scientific and academic authors when it comes to AI do not necessarily align with purely commercial authors. Public-minded authors wish for their works, or are required by the public funding, to be widely used as possible in the public interest and to support technological and scientific advancement. This tension between these different aims is another reason why targeted cultural policies and not copyright revision is required to support commercial creators.

Question 8. Do you agree that rights should be reserved in machine-readable formats? Where possible, please indicate what you anticipate the cost of introducing and/or complying with a rights reservation in machine-readable format would be.

[5] Amanda Levendowski, 'How copyright law can fix artificial intelligence's implicit bias problem' (2018) Washington Law Review 93

We consider a reservation of rights to be neither desirable nor practical. When copying is undertaken for the purpose of analysing trends, patterns, facts, and other non-protected elements of a work, technical copying to access content in the public domain should be permissible.

It is highly unlikely that university staff, researchers, knowledge transfer collaborations or even medium sized businesses will have the necessary resources to comply with rights reservations, whether expressed through technical measures or contractual terms and conditions on the internet, as highlighted in the *Kneschke v. LAION* ruling of the Hamburg Courts.[6]

From a public policy perspective, imposing such burdens—particularly when competitors outside of Europe are not subject to similar constraints—appears counterproductive. Such measures risk either stopping activity entirely or driving machine learning research and development abroad.

We understand that the Commission is already deeply concerned by the level of opt-outs that have been put in place by right holders.

The addition of reservation of rights would degrade current research resources. For example, universities and other educational organisations, SMEs, university spin-outs, even medium sized companies make significant use of Common crawl, a database of over 250 billion webpages. A key use-case of Common crawl is to test the accuracy of machine learning models on real data.

Given the lack of standards and the ruling in *Kneschke v. LAION* we find it almost inconceivable that a European version of Common Crawl could be created. Even if the practical issues were overcome, the content would be neither as wide nor as deep. Creating a smaller, less rich dataset will likely promote bias and poor predictive capabilities for those models that use it.

C.4 Transparency

Question 17. Do you agree that AI developers should disclose the sources of their training material?

Yes. The government should encourage the development and adoption of **best practice standards for transparency industry-wide**, so that AI developers take a consistent approach to citing their sources wherever possible.

However, its practical implementation presents significant challenges.

---

6 https://www.euipo.europa.eu/en/law/recent-case-law/germany-hamburg-district-court-310-o-22723-laion-v-robert-kneschke

A key concern is that, given the prevalence of techniques such as transfer learning and model distillation, many organisations may lack precise knowledge of the sources of their training data.

Moreover, we do not support the use of transparency requirements as a mechanism for restricting machine learning research within academic institutions. There is already evidence suggesting that, under the EU AI Act and Article 4 of the CDSM Directive, publishers may leverage transparency obligations to scrutinise the content utilised under Article 3 of the Directive, particularly in cases where research organisations engage in public-private partnerships. We contend that transparency measures must be implemented with careful consideration to avoid stifling upstream innovation within universities.

C.5 Wider clarification of copyright law

Question 26. Does the temporary copies exception require clarification in relation to AI training?

Yes. **S28A CDPA should be broadened to clarify that all acts of extracting information from copyright works are outside the scope of copyright.** Over the years, due to the development of copy-dependent technologies such as computers, the reproduction right has expanded far beyond the original scope and purpose of copyright. As the European Copyright Society recently reminded us, the 'policy choice of including any technical, even if fugitive, fixation of a work within the scope of reproduction right, made by the EU lawmaker as early as the 1991 directive on computer programs, could have been different and remains challenged by several copyright scholars.'[7] The UK government now has the opportunity to bring copyright back to its original goal: to protect against substantial similarity that can be perceived by a human. To do so, it can take inspiration from other jurisdictions with strong AI and creative industries, such as Japan's Art 30-4.[8] We suggest a possible approach under the next question.

Question 27. If so, how could this be done in a way that does not undermine the intended purpose of this exception?

The purpose of the exception is to allow mere processing of information by machines and therefore our suggested amendments directly support its aim more than the existing EU formulation does. S28A was implemented as a mandatory exception under the Information Society Directive (2001/29/EC). Now, it can and should be broadened to reflect the current technological landscape and to capture future changes in technology. In our view, the

[7] Copyright and Generative AI: Opinion of the European Copyright Society (January 2025). Available at: https://www.create.ac.uk/blog/2025/02/07/opinion-by-the-european-copyright-society-on-generative-ai/

[8] For an overview of different approaches to these issues, see Mendis, Dinusha and White, Benjamin and Hong, Dukki, Copyright and Open Norms in Seven Jurisdictions: Benefits, Challenges & Policy Recommendations (February 15, 2024). https://ssrn.com/abstract=4728782

following text would achieve this future-proof technological neutrality while also providing sufficient clarity to beneficiaries:

28A Extracting information and other technical uses

Copyright is not infringed by the copying of works which is transient or incidental, is an integral and essential part of a technological process, is aimed at extracting information from the work, or by any other technical use of the work that does not involve or cause a human audience to enjoy the work.

C.6 Encouraging research and innovation

Question 28. Does the existing data mining exception for non-commercial research remain fit for purpose?

No. The existing exception for text and data mining does not reflect current research practice. An expanded text and data mining exception, combined with an open norm and an exception for all technical uses (see above), would strike the right balance between keeping the law flexible and future-proof, and providing clarity to beneficiaries of exceptions. The key issues of the existing s29A CDPA are as follows:

- By only covering the 'making' of copies [s29A(1)] for non-commercial research purposes and prohibiting the unauthorised transfer of such copies [s29A(2)(a)], the exception does not allow knowledge transfer in partnerships between different academic institutions, let alone in public-private partnerships. Without authorisation, datasets obtained or created under the existing exception cannot be shared lawfully with the wider research community.
- Although in principle the exception cannot be overridden by contract [s29A(5)], in practice for various reasons contractual arrangements often regulate access to and the computational use of data for research. The 'lawful access' requirement is ambiguous and open to conservative interpretations, leading to risk averse behaviour and rights holders' decisions shaping data-driven research.
- The relationship between technological protection measures (TPMs) and exceptions, including the TDM exception, remains problematic.[9]
- Copyright law is not the only relevant legal regime in relation to data scraping and mining activities. The sui generis database right, data protection law, contract law, confidentiality, and competition law pose additional risks and uncertainties. Some of these should be addressed in the context of this consultation.

Now that the UK is not bound to the artificial and outdated distinction between commercial and non-commercial purposes inherited from EU law, we recommend the following

[9] With Arcadia funding provided by Knowledge Rights 21, CREATe has recently produced evidence in this area. See Erickson, E. et al. (2024). Evidence on Technological Protection Measures: impact on research, education and preservation. Available at: https://www.create.ac.uk/project/public-domain/2024/06/16/evidence-on-technological-protection-measures-impact-on-research-education-and-preservation/

amendments to s29A to address the issues above:

- The ‘non-commercial’ purpose restriction should be removed. It creates uncertainty, prevents public-private partnerships, and conflicts with the common law tradition.
- The exception should remain available to everyone, and focus on a broader research purpose that can capture how R&D will evolve in the future.
- The exception should clearly permit digitisation for the purposes of the exception and their preservation.
- Sharing with others and retaining copies made under the exception should be allowed to the extent justified by the permitted purpose. This is to reflect current research practice, including knowledge transfer, public-private partnerships, peer-review, accuracy checking and experimentation.[10]
- The exception should remain non-overridable by contract [s29A(5)].
- The exception should apply also to the sui generis database right to reduce uncertainties around its application.
- A more streamlined mechanism for the exception’s beneficiaries to remove Technological Protection Measures (TPMs) is needed to unlock data across the UK research infrastructure. We recommend following countries like Slovenia who require rightsholders give access in 72 hours. Where this is not given, as in many other jurisdictions, given that a permitted act is lawful, unilateral circumvention should be allowed.[11]

Question 29. Should copyright rules relating to AI consider factors such as the purpose of an AI model, or the size of an AI firm?

AI models vary in their purpose, and this variability is not sufficiently factored into the government’s proposed approach. The issues outlined in the consultation appear to be predominantly driven by concerns related to **generative AI models**, particularly those deployed within the entertainment industry. In contrast, there is minimal—if any—consideration given to the application of AI across STEM industries, the NHS, public-private partnerships, and other critical sectors, which may not require training, and may not be generative - i.e. “non-expressive” models. As an example, in the United States such models are protected under fair use principles.[12]

---

[10] In its current form, s29A(2)(a) stating that copyright is infringed when a copy made under the exception is ‘transferred to any other person’ is too restrictive. Combined with a conservative interpretation of the ‘lawful access’ requirement, it poses major obstacles in research partnerships, such as the Living with Machines example mentioned above.

[11] 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),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.

[12] *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630. *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 8. *Authors Guild v. Google,* 804 F.3d

D.2 Policy Options

Question 30. Are you in favour of maintaining current protection for computer-generated works? If yes, please explain whether and how you currently rely on this provision.

No. We agree with the assessment given above that the protection granted to computer-generated works (CGW) in the UK is unlikely to have benefitted the development of AI technology, or indeed, encouraged its use. Moreover, it is clearly non-compliant with the Berne Convention which only protects works produced by a human author. There is little to no evidence that AI innovation is driven by the prospect of licensing CGW, and we don't see the logic in protecting outputs by users of these tools (unless there is substantial human creativity and effort involved, for example for "AI assisted" works). In the interests of streamlining the UK's copyright framework, **we believe s9(3) should be removed along with any other related provision such as s12(7)**.

The originality threshold of a work as an author's own intellectual creation has been confirmed by the 2023 *THJ vs Sheridan* ruling: "What is required is that the author was able to express their creative abilities in the production of the work by making free and creative choices so as to stamp the work created with their personal touch…This criterion is not satisfied where the content of the work is dictated by technical considerations, rules or other constraints which leave no room for creative freedom".[13] Refer to the CREATe response to this consultation for more information on the applicability of this case to CGWs.

AI-generated content is being built into our day-to-day use of digital tools. Users of Copilot, an AI-based tool embedded in Microsoft Office products, may generate new content by typing in prompts to enhance productivity (e.g. summaries of text, automated note-taking, simplified language versions). In another example, users may ask ChatGPT to create essays based on a particular topic, or create an image based on any number of inputs. The AI outputs may or may not go through a process of refinement by the user. In many cases, the level of human effort required is tiny. As currently written, the definition of CGW would include this type of output, despite the fact that such applications are not intended to be used to generate original works, as defined by the CDPA. Works substantially created by a human but assisted by digital tools including AI-based tools should be defined as literary, dramatic and musical works, artistic works, etc. and receive the **usual protections for those types of works**.

[13] https://www.bailii.org/ew/cases/EWCA/Civ/2023/1354.html

D.5 AI output labelling

Question 40. Do you agree that generative AI outputs should be labelled as AI generated? If so, what is a proportionate approach, and is regulation required?

Yes, where necessary. Best practice standards should be agreed and implemented with government support to ensure that publicly-accessible genAI outputs are labelled accurately as such to mitigate the proliferation of misinformation and the risk of harm to the public and other users. The EU AI Act rightly recognises the importance of "reliable, interoperable, effective and robust" labelling and captioning to improve the effectiveness of this practice.[14] Any best practices developed in the UK should do likewise. The AI Act includes steep financial penalties for non-compliance with labelling obligations (a fine of up to 15 million EUR or up to 3% of total global annual turnover). Reasonable and proportionate fines should be considered as part of the UK's AI Safety Institute research factoring in the different contexts and risk of harm to the public.

D.6 Digital replicas and other issues

Question 43. To what extent would the approach(es) outlined in the first part of this consultation, in relation to transparency and text and data mining, provide individuals with sufficient control over the use of their image and voice in AI outputs?

This question pertains solely to generative models, where there is a potential for the outputs to be substantially similar to the inputs. However, as previously noted, many AI models are non-expressive, meaning this question is not relevant to the majority of AI models. Existing law already provides remedies for right holders in such cases. Consequently, we believe there is no need for additional legal intervention to address concerns about overfitting in AI models. Furthermore, image and voice are potentially protected by a wide body of law ranging from data protection, licensing to tort law – it goes without saying these all sit beyond copyright law.

[14] https://www.euaiact.com/article/50