Ref. Ares(2026)6439675 - 25/06/2026

ekip Response to the Targeted initiative for a better copyright environment for European creativity and innovation: Call for evidence for an impact assessment

June 24, 2026

To: European Commission, Directorate-General for Communications Networks, Content and Technology (DG CNECT), Unit I.2 (Copyright)

From: *ekip,* the European Culture and Creative Industries Innovation Policy Platform (EU Grant agreement: 101112111)

Subject: Critical considerations concerning the cultural heritage sector and creative industries regarding works in copyright on the open web

ekip is an EU Horizon-funded research project and collaborative platform designed to boost the development of innovation capacity in ecosystems to include the Cultural and Creative Industries (CCI) (see https://ekipengine.eu). We thank the European Commission for providing us this opportunity to bring awareness to two sectoral areas concerning updates of legislation: 1) cultural heritage and GLAM (galleries, libraries, archives and museums) institutions, and 2) creative industries. We provide the following evidence in summary form.

This report is based on ongoing research by *ekip* into the role of cultural heritage, GLAM institutions, and the creative industries in the digital commons. This research explores how the digital commons removes barriers to access and reuse of digital heritage content, enabling institutions and creators to make their work available to large audiences.

Submission prepared by:

Morgan Currie, Senior Lecturer; Science, Technology and Innovation Studies; ekip; University of Edinburgh

Caitlin McDonald, Research Assistant, ekip, University of Edinburgh

Nicola Osborne, Manager of the Institute for Design Informatics; ekip; and BRAID Creative Industries Lead, University of Edinburgh.

Matea Senkic, Research Assistant, ekip, University of Edinburgh

ekip Response to the Targeted initiative for a better copyright environment for European creativity and innovation: Call for evidence for an impact assessment

Anna Talley, Research Assistant, ekip, University of Edinburgh

Jennifer Volk, Research Assistant, ekip, University of Edinburgh

High-level summary

The EU's *cultural heritage and GLAM (galleries, libraries, archives and museums) institutions* strongly promote protecting copyright, but copyright restrictions are at odds with the AI sector's approach of indiscriminately scraping the open web. The cultural heritage sector is, therefore, increasingly concerned about 1) AI scraping of copyrighted content in their collections, 2) the financial costs resulting from AI scraping on some institutions' core IT infrastructure, and 3) new enclosures of digital collections as rights holders in the sector withhold, remove or block content from the open web due to copyright concerns.

Our response also notes that the *culture and creative industries* (CCIs) currently face asymmetric relations with the AI industry regarding the development and deployment of AI systems. Current laws do not support CCIs to uphold access restrictions on their copyrighted works by AI systems, and content creators lack channels to understand when their work is used by AI companies or receive potential recompense for their contribution.

We propose four approaches forward that would strengthen the positions of copyright holders in the culture and creative industries and cultural heritage and GLAM sectors:

1. Clarify areas of ambiguity in current copyright law to **reduce burdens on rights holders to reserve in-copyright works** from commercial AI scraping.
2. Create and promote **authenticity standards** that support companies to adhere to open licenses.
3. Develop easy-to-use (including by creative small and medium-sized enterprises (SMEs) and individual SMEs) **opt-in registries** for content creators and stewards of creative content that are shared with AI companies.
4. Facilitate **transparency by AI companies** by creating channels by which rights holders can query commercial use of material.

The onus should be on AI developers to ensure copyrights are upheld and seek permissions – and ideally offer recompense - to use data used to train their models.

1) Cultural heritage, GLAM and copyright

The EU's cultural heritage and GLAM sectors strongly promote protecting copyright while making pathways to open access and open knowledge on the open web.

Over the past several decades, the European Union has supported a set of measures for digitising cultural heritage and making it accessible online (EU Commission, 2011; Cousins, 2018; Ziku and Fabos, 2022). EU policy strongly recommends that knowledge, information and heritage be shared in an open way, using open access licenses wherever possible, while also recognising the importance of respecting rightsholders' interests (UNESCO, 2015; Council of Europe, 2005; European Parliament, 2019).

Cultural heritage institutions now use a range of licenses for their holdings; most place restrictions on reuse. Open access licenses, for instance, should not be confused as being copyright free. Open licenses, such as Creative Commons, make content available for certain types of public reuse while retaining it as copyrighted material. Many open access works have copyright allowing reuse, but only if the originator is attributed (e.g. Creative Commons' 'Attribution (CC-BY)'), or only for non-commercial purposes (e.g. Creative Commons' 'Attribution-NonCommercial (CC-BY-NC)'). Such rights restrictions are at odds with the AI sector's approach of indiscriminately scraping the open web – that is, using bots (software programs) that crawl the internet and download and format web data for training AI models.

In addition to signaling reuse restrictions through licenses, institutions with websites on the open web can restrict bots from scraping by using 'do not crawl' standards, such as robots.txt, watermarking and Terms & Conditions, to signal to bots that their site, or parts of their site, does not permit access. However, these current technical standards and licensing are not universally respected or enforceable. Our research found – corroborating other accounts (Weinberg, 2005) – that these standards are being ignored by the majority of AI scraping bots.

Generative AI output created with copyrighted works, but without lawful permission, is copyright infringement according to Articles 5(a) and 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of the DSM Directive. The Text and Data Mining (TDM) legislation in the EU Copyright Act (2019) requires rightsholders to "expressly reserve" their works (i.e. to opt out) from TDM (article 4) unless AI training is for research.

Currently, there are no adequate industry standards for legally opting out of commercial AI TDM (e.g. scraping) or upholding copyright restrictions.

The Copyright Act (2019) should be modified to make these regulations stronger against copyright violations. Currently TDM is lawful unless a content holder reserves their rights, but the legislation offers no single and undisputed mechanism to reserve such rights, placing unnecessary burdens on rights holders. Legislation should provide clear definitions of acceptable and legally enforceable rights reservations and support the design of standards for opting out of bot scraping, as well as provide clear definitions between TDM for research versus commercial exploitation.

EU Copyright could go further, moving to an opt-in model that requires permissioned material for training data models to reward businesses that respect the copyrights of high-quality content stewarded by cultural heritage and GLAM sectors.

Cultural heritage and GLAM institutions are experiencing scraping of copyrighted content on the open web and fear bots' growing capacities to access restricted material.

Use of copyrighted content to train AI is rife and a well-known, much-debated concern for copyright holders (as evidenced in numerous court cases across the world, such as the Authors Guild et al. v. OpenAI Inc. (2023–Present)). Cultural heritage and GLAM institutions that allow open access to their collections also share this concern. In our research, we are finding that institutions are currently experiencing heavy bot activity presumed to come from AI companies; these bots do not differentiate across content with different licenses. Bot scraping for AI models tends to be unselective, large scale, and for commercial, rather than public, ends.

Our emergent research finds that people in the sector are concerned that AI bots will develop in sophistication enough to access material that is not on the open web but behind logins or firewalls. In addition to concerns for rights holders, interviewees from the GLAM sector (in a variety of roles) have expressed worries that bots will soon be able to access digital objects related to sensitive bioanthropology objects, such as biomedical data about human remains, and images of archeological sites in restricted locations.

The sector is widely concerned about AI scraping of copyrighted content and the financial costs resulting from this, along with new enclosures of digital collections reshaping the open web.

Our research finds evidence that, while many in the sector see enormous potential for AI scraping in terms of resource discovery for heritage collections, several are concerned about AI scraping of copyrighted materials and the wider implications this has for the cultural heritage and GLAM sectors. We find two main implications for the sector that bear on this Response: 1) new enclosures of heritage materials and open knowledge and 2) financial burdens on cultural institutions of high-volume, consistent scraping.

People in the sector we spoke to offered examples of how AI scraping is leading to new enclosures of knowledge and heritage material as collections shut off access to bots or creators refrain from depositing content due to copyright concerns. For example, our research found that some potential contributors to collections and repositories, such as researchers, are not depositing content into open collections if they believe their content will be scraped for commercial purposes.

Compounding these concerns are the additional monetary costs due to heavy bot traffic straining servers, impacting the day-to-day operating costs of institutions; as one institution told us, "we were not built to handle 735 million search requests a month." In our research, this cost most often takes the form of additional labour hours required for web developers to mitigate overburdened servers. This trend undermines cultural heritage institutions that hold high-quality content but, in some cases, must now shoulder new financial costs without additional recompense.

As a response to these two concerns, institutions are adopting software intermediaries such as Cloudflare to prohibit bot activity from certain geographies or certain activities on their website, such as search, or disallowing bots altogether; such activity creates new firewalls that are currently curtailing the open web.

One result of bot blocking is that high-quality, diverse, national and European heritage content is less accessible to AI companies, a loss which could have important implications for economic growth in the AI sector.

Reactive enclosure by blocking bot activity to protect intellectual property rights and mitigate infrastructural costs also leads to negative effects on public interest usages of bots for research, art and web preservation. Web preservation – including for legal deposit libraries – has become much more complicated in recent months as web pages become inaccessible to crawling due to installing cybersecurity software that blocks bots.

2) The Culture and Creative Industries and copyright

The Culture and Creative industries (CCIs) face asymmetric power relations in the development and deployment of AI systems.

AI has been mostly developed and applied by AI labs run by large tech companies and online platforms, tech startups, and specialised research centres. Large tech platforms and AI developers have greater access to data, computing power, financial resources, technical expertise, and market reach. As a result, smaller players such as independent content creators, artists, and small and medium-sized enterprises (SMEs) within the CCIs face reduced influence and heightened dependence (ekip Policy lab, 2025).

Industry surveys report that creative professionals are already experiencing reductions in commissions, audience traffic, attribution and revenue through generative AI creations, based on training on copyrighted works. For example, the Association of Illustrators reported in 2025 that 32.4% of illustrators had lost work to AI alternatives (2025). Creators are concerned that AI systems may derive economic value from copyrighted content while providing limited recognition and remuneration to individuals whose works contribute to those systems. The cultural sector thus becomes poorer and less diverse due to a lack of support for original work.

Our research suggests that lowering or removing copyright protections to facilitate AI development and innovation would primarily benefit the business models and data acquisition practices of the largest, best funded Generative AI/LLM corporations rather than smaller creators (Osborne et al., 2025). In contrast, protecting copyright can support innovation while encouraging more responsible and ethical use of others' data and creations.

The Culture and Creative Industries experience uncertainty regarding transparency around use of their work by AI and upholding restrictions of their copyrighted works in AI systems.

The EU copyright legislation (EU 2019/790) and the EU AI Act set the regulatory framework on text and data mining (TDM). However, as mentioned above, practical implementation challenges remain, particularly for cultural content used in AI training. The first challenge concerns transparency around AI training data. Creative and cultural actors lack the ability to query how their work is accessed, reused, or incorporated into AI training.

The second challenge concerns copyright reinforcement. As mentioned, existing mechanisms, including robots.txt files, website terms and conditions, metadata, watermarking and licensing notices, are inconsistently recognised and frequently ineffective in preventing the use of content in AI training datasets (Kretschmer et al., 2025; Lee et al., 2026). Current EU copyright legislation does not offer details on exact implementation of reservation or opt-out in the context of AI integration. As a consequence, many creators face a disproportionate burden in monitoring or challenging the use of their work in AI training. This burden falls on individual creators and SMEs who often lack the legal, technical, and financial resources required to track the use of their content and so have a very high barrier to countermand AI training based on their work once it has occurred (Osborne et al., 2025).

Our research has also found the effects of this vagueness on creators wishing to deploy AI themselves. Creators, such as those in the music industry, have expressed reservations about employing AI tools; their skepticism is rooted in the legal ambiguities surrounding the interaction with intellectual property (ekip Policy lab, 2025). This ambiguity includes unclear and varied copyright status (including copyright protection) of works created by or with/in collaboration with AI, particularly generative AI (Erickson 2024, Kartunnen 2025), which impacts the potential for (ethical) exploitation and business models around creative IP.

Unequal access to data, expertise and resources risks widening existing inequalities and limiting participation of CCIs in AI development processes.

Our research highlights that access to AI-relevant datasets and computing infrastructure is uneven across sectors. Smaller creative actors lack resources to access diverse datasets which can hinder the development of inclusive and culturally relevant AI solutions. SMEs and independent creators often lack the time, funding and organisational capacity to build or acquire the necessary expertise. Furthermore, the landscape is complicated by AI

startups that lack specialised knowledge tailored to the CCIs, thus not enabling these small entities to leverage AI to their advantage. This gap in domain-specific understanding by AI startups creates a barrier to effectively serving the unique needs of independent artists and SMEs within the CCIs.

As a result, CCIs are often excluded from AI development processes and consulted only at later stages. Decisions regarding the use of copyrighted content in AI systems are frequently made without meaningful participation from creators, resulting in systems that do not adequately reflect the cultural values of CCIs or their specific concerns.

Recommendations

The EU, which has historically been an international leader in designing regulation of the tech industry, could take a guiding role by strengthening the positions of copyright holders in the creative industries and cultural heritage and GLAM sectors. The EU (and its member states) should play a role in regulating copyrighted content created through public heritage institutions funded by taxpayers through national and local government. The government should likewise support Europe's thriving creative industries and artists by giving their needs parity with the AI industry. By way of context, the cultural and creative ecosystem represents 3.95% of added value for the EU economy and employs 8.02 million people and is therefore of comparable scale and impact to the digital sector – which provides 5.17% of added value, employing 6.8 million people (European Commission Staff 2024) - and should therefore be afforded similar support in policy.

We see four approaches forward:

1. Clarify legal areas of ambiguity in current copyright legislation to allow people and institutions to opt out of TDM.

The current ambiguities in the 2019 Copyright in the Digital Single Market Directive have led to cases of complainants who do not succeed in proving breach of copyright because the burden of proof is set too high, requiring copyright holders to anticipate any and all use of their material for TDM in advance.

The relevant example is of Amsterdam District Court - DPG Media v. HowardsHome; here the defendant had scraped news from EU-based publishers, who were suing for breaching

terms of bans against mainstream bots in their robots.txt file. Because HowardsHome was not explicitly banned in the terms, this did not count as "expressly reserving" rights - demonstrating the inadequacy of standards for opting out of AI scraping for commercial ends.

2. Develop, support and regulate authenticity standards

We recommend the EU Commission regulate emerging authenticity standards for machine-readable traceability and provenance that automate crediting authorship. Such a standard enables vastly enhances transparency and affords key opportunities for a more ethical and economically fairer use of data in the training of AI. For instance, it would enable copyright holders to properly exercise their rights, including the right to seek recompense for their works/intellectual property (IP); it would enable rights and collections holders in the GLAM and creative sectors to ensure adherence to 'Attribution CC-BY' open licenses.

Current technical standards for provenance and attribution include the 'Content Credentials', as part of the Coalition for Content Provenance and Authenticity (C2PA) (Collomosse, 2024). The Ownership, Rights, and Attribution (ORA) framework enables rights management for digital content through tokenised licensing and distributed ledger technologies (DLTs) that work with C2PA (Elsden et al, 2026). ORA provides provenance information recorded on a DLT, along with any changes and edits made to a piece of content. The Commission could thus support the creation and upholding of a European authenticity standard enabling fairer and more scalable approaches to traceability in support of rights holders.

3. Support an opt-in registry and related infrastructures for collaborations between AI companies and the cultural heritage sector.

Whilst the EU AI Act currently proposes an Opt Out approach for data holders, we have noted above the lack of current solutions and practical challenges of this approach – which is particularly impractical for individual creators and smaller scale rights holders (noting that the creative and cultural industries are composed of 99.9% SMEs in Europe (European Commission Staff 2021)). Additionally, participants in our research described a lack of consideration by tech companies towards the negative impact that scraping has on copyright protections and core IT infrastructure. One way forward is to prohibit blanket

scraping that disregards copyright and instead permit ingestion of explicitly permitted works in copyright for AI training, supported by an opt-in registry. This registry would provide a source of training data with clear permissions over the kinds of use that could be made of this data.

This recommendation seeks to find a more positive way forward for rights holders and AI companies to work together. For instance, experts in our study proposed a mutually beneficial consortium between cultural heritage institutions and 'good actor' AI companies within a technical environment that allows scraping without disruption and with controls around copyright restrictions. "Any AI company that wants to play by the rules is to come inside the gates, play by the rules, get access" as one interviewee suggested. This approach also benefits from collective representation by the sector, rather than individualised one-to-one negotiations with tech companies.

4. Support transparency by AI companies regarding copyrighted material in their training data and models

This recommendation is in line with the EU AI Act 2024/1689 transparency requirements requiring AI developers to disclose information on the model to prevent it from generating illegal content and to publish summaries of copyrighted data used for training. The EU should move to align copyright law with these regulatory requirements of AI developers. Copyright holders should be able to understand when their work is used and be able to exercise their rights over their protected content, to ensure the ethical and legal use of copyrighted material in AI. Such measures also comply with individuals having the capacity to request deletion of their data under GDPR.

In addition to ensuring copyright holders are aware of how their data is being used, this measure will also enable potential recompense for creators, ensuring that they benefit from value being derived from their work rather than AI companies being the sole financial beneficiary from exploitation of others' IP.

References

Association of Illustrators (2025) *UK Illustrators draw the line: AOI publishes AI survey results*. Available at: https://theaoi.com/news/the-aoi-ai-survey-results

J. Collomosse and A. Parsons, "To Authenticity, and Beyond! Building Safe and Fair Generative AI Upon the Three Pillars of Provenance," in IEEE Computer Graphics and Applications, vol. 44, no. 3, pp. 82-90, May-June 2024, doi: 10.1109/MCG.2024.3380168.

Council of Europe. (2005). Convention on the Value of Cultural Heritage for Society (Faro Convention, 2005). https://www.coe.int/en/web/culture-and-heritage/faro-convention

Cousins, Jill (2018). Building a commons for digital cultural heritage. SMK – National Gallery of Denmark in Copenhagen (Statens Museum for Kunst). Available at: https://www.smk.dk/en/article/building-a-commons-for-digital-cultural-heritage/ (accessed 1 June 2026).

ekip Policy Lab (2025) *Artificial Intelligence for the Culture and Creative Industries – Policy Recommendations from Policy Lab 1*. July 2025. Available at: https://ekipengine.eu/wp-content/uploads/2024/11/Policy-Lab-1-AI-and-CCI-1.pdf

The European Parliament and the Council of the European Union (2019). Directive (EU) 2019/1024 of the European Parliament and the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information, EU 2019/1024. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1561563110433&uri=CELEX:32019L1024

Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market (DSM Copyright Directive) (2019). Available at: https://eur-lex.europa.eu/eli/dir/2019/790/oj

Elsden, C., Liddell, F., Moruzzi, C., Schafer, B., Tallyn, E., Morgan, E., Dixon, B., Balan, K., & Collomosse, J. (2026). Building trust and developing value between creative rightsholders and genAI through tokenised licensing and content authenticity tools. The Paris Journal on AI & Digital Ethics, 126-139. https://doi.org/10.65701/j7d2m0q9t5

Erickson, K. 2024. Copyright protection in AI-generated works: Evolving approaches in the EU and China. Creative Industries Policy and Evidence Centre blog, 17 July 2024. https://pec.ac.uk/blog_entries/copyright-protection-in-ai-generated-works-2/

European Commission 2024. Cultural and Creative Industries. European Commission Internal Market, Industry, Entrepreneurship and SMEs: Cultural and Creative Industries [website]. https://single-market-economy.ec.europa.eu/sectors/cultural-and-creative-industries_en

European Commission Staff 2021. Annual Single Market Report 2021 Accompanying the Communication from the Commission to the European Parliament, the Council, the European

Economic and Social Committee and the Committee of the Regions Updating the 2020 New Industrial Strategy: Building a stronger Single Market for Europe's recovery [final working document SWD/2021/351, document number 52021SC0351]. EUR-LEX. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52021SC0351

Kartunnen, S. 2025. Copyright of AI-generated works: Approaches in the EU and beyond. EPRS: European Parliamentary Research Service (PE 782.585), December 2025. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)782585

Kretschmer, M., Meletti, B., Bently, L., Cifrodelli, G., Eben, M., Erickson, K. et al. (2025) 'Copyright and AI in the UK: Opting-In or Opting-Out?', *GRUR International*, 74(11), pp. 1055–1070.

Lee, C.P., Hong, R., Jiang, H.H., Plotnik, A., Agnew, W. and Morgenstern, J.H. (2026) 'How Do Data Owners Say No? A Case Study of Data Consent Mechanisms in Web-Scraped Vision-Language AI Training Datasets', *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(45), pp. 38808–38816. https://doi.org/10.1609/aaai.v40i45.41225

Osborne, N., Parkinson, C.A., Terras, M. and Cruz, J. (2025) *Edinburgh College of Art, University of Edinburgh Response to the UK Government Intellectual Property Office Consultation: Copyright and Artificial Intelligence*. Available at: https://doi.org/10.5281/zenodo.15002867

Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (2024) Official Journal L 2024/1689. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

UNESCO (2015) Recommendation concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form. UNESCO. Available at: https://www.unesco.org/en/legal-affairs/recommendation-concerning-preservation-and-access-documentary-heritage-including-digital-form (accessed 27 April 2026).

Weinberg M (2025) Are AI Bots Knocking Cultural Heritage Offline? Glam e-lab, June. NYU: Engelberg Center on Innovation, Law and Policy. Available at: https://www.glamelab.org/products/are-ai-bots-knocking-cultural-heritage-offline/ (accessed 20 April 2026).

Ziku M and Fabos B (2022) Digital Community Heritage and Open Access. CC Open Culture Working Group Digital Community Heritage, 30 November. Creative Commons. Available at: https://cc-digital-community-heritage.pubpub.org/pub/report (accessed 27 April 2026).