Artificial Intelligence for Cultural Heritage Research: the Challenges in UK Copyright Law and Policy

Dr Paula Westenberger*, Dr Despoina Farmaki[1]

Abstract

Artificial intelligence (AI) is revolutionising our relationship with cultural heritage, enhancing access to, engagement with and preservation of collections and heritage sites. AI is also being used as a valuable research tool in the context of heritage collections. However, as materials protected by copyright may be used in AI development, training and use, copyright law can become an obstacle to important AI deployments in the heritage sector, an area which is currently understudied from the United Kingdom (UK) perspective. This article explores the intricate interplay between cultural heritage, AI and copyright law, demonstrating the main copyright law and policy challenges facing cultural heritage professionals and researchers in using AI in the UK for heritage research. It highlights the complexity and uncertainties as regards the current Text and Data Mining exception in the UK Copyright Designs and Patents Act, emphasising the need for an improved legal framework that balances copyright protection with the benefits of AI for cultural heritage research and management. It also reveals the underrepresentation of the heritage sector in AI regulation and copyright policy discussions in the UK. This exploration underscores the imperative for an inclusive policy dialogue that considers the perspectives and evidence of the cultural heritage sector in its full breadth and diversity (including related researchers) in shaping copyright law reform and AI regulation, and for further research to be carried out in this field.

Keywords: cultural heritage, research, artificial intelligence, copyright, text and data mining, United Kingdom

---

Table of Contents

INTRODUCTION

The adoption of new technologies in the cultural heritage sector often raises issues around copyright law, particularly whether current legislation is fit for purpose. Historically, copyright law has constantly adapted to address technological advances (Gervais *et al*, 2024, p. 28). With the increasing use of Artificial Intelligence (AI) across sectors, copyright laws are once again under scrutiny. However, there is a specific need to address the unique copyright challenges AI presents to the UK heritage sector, which are currently underexplored.

Cultural heritage institutions, entrusted with the preservation of the legacies of humankind, are navigating complex legal terrain in their pursuit of modernisation. AI is revolutionising the forms of usage of cultural heritage materials, including as a tool for preservation, research, and dissemination. The volumes of data now available create traction to digital humanities approaches, facilitating the processing of data in scale "to yield new analytical insights that were not possible at the level of individual documents and sources" (Ahnert *et al.* 2023). While the importance of digital heritage collections as AI data[2] is being discussed in current academic

---

[2] On the concept of "collections as data": Cory Lampert and Emily Lapworth, 'What do we mean by "Collections as Data" (CAD)?' https://www.library.unlv.edu/whats-new-special-collections/2020/2020-03/what-do-we-mean-collections-data-cad-cory-lampert-emily; and Mia Ridge, 'Keynote video "Evolutionary Innovations: Collections as Data in the AI era" for Making Meaning 2024' https://www.miaridge.com/keynote-evolutionary-innovations-collections-as-data-in-the-ai-era-for-making-meaning-2024/ [Accessed February 19, 2025].

and sector-specific debates, there is scope for investigation of related copyright challenges in the UK.[3] Since copyright materials may be used in AI development, training and use, the current issues and uncertainties in copyright law may pose obstacles and ultimately hinder important AI applications in the traditionally risk-averse heritage sector. Although these issues are being explored in Europe,[4] the UK lags behind, making this a crucial and novel area of investigation in a jurisdiction that can influence legislations worldwide. This paper therefore aims to contribute to current underdeveloped UK policy discussions on AI and copyright with such a heritage perspective, by analysing the particular shortcomings of, and proposing solutions for, the UK Text and Data Mining (TDM) exception for non-commercial research, which we argue is not currently fit for purpose when applied specifically to heritage research.

Section 1, highlights the importance of AI in the heritage sector with UK examples. Section 2 examines key issues in UK copyright law, particularly the TDM exception, crucial for AI development and knowledge discovery in heritage research. Section 3 addresses the state of current policy discussions on copyright and AI regulation, emphasising the need for greater involvement and evidence from the cultural heritage sector.

1. ARTIFICIAL INTELLIGENCE AND CULTURAL HERITAGE

AI refers to computer programmes that perform tasks associated with human intelligence, such as language comprehension, image recognition, and learning from experience (Pavis, 2023). AI is the umbrella term including machine learning, whereby a computer programme is taught to identify patterns in data and apply this knowledge to new data (Drexl *et al.,* 2019; Iglesias Portela *et al*., 2019).

With libraries, archives and museum collections increasingly digitised, AI is revolutionising heritage practice and research. Pavis (2023) identifies three overlapping heritage areas of AI application when used legally and ethically: "heritage and collections management, use and research; visitor experience; and general business operations and management". AI enhances

---

[3] Noting that relatively few books and articles address the practical steps in getting hold of data, and the restrictions that come with it, discussing specifically copyright and contractual challenges Ahnert *et al.* (2023) p. 23.

[4] Europeana, 'Copyright and Digital Cultural Heritage' <https://pro.europeana.eu/page/copyright-and-digital-cultural-heritage>; Magdalena Pasikowska-Schnass and Young-Shin Lim. 'European Parliament Briefing: Artificial intelligence in the context of cultural heritage and museums' (PE 747.120 – May 2023) https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747120/EPRS_BRI(2023)747120_EN.pdf [Accessed April 30, 2024].

engagement with cultural heritage collections by producing innovative documenting, managing, and visiting tools (Bordoni *et al*., 2016). Tools and techniques of AI also "make it possible to build the fine instruments that augment the day-to-day work of librarians and the researchers they serve" (Coleman, 2019). Digital humanities scholar Professor Jane Winters explains that "AI is essential for cleaning, exploring, and visualizing archival and special collections, especially with born-digital archives".[5] According to the European Regions Research and Innovation Network (ERRIN), "AI is set to revolutionise cultural heritage by enhancing the preservation, restoration, and accessibility of artefacts and historical sites", aiding conservation efforts by detecting early signs of deterioration.[6] Recent advances in AI have led to innovative heritage applications e.g. in archaeology (Chetouani *et al*., 2020; Ostertag and Beurton-Aimar, 2020); document digitalization and character recognition (Nguyen *et al.*, 2020); discovery, description, classification, and preservation (Girbacia, 2024); and reconstruction of heritage buildings and sites (Arzomand *et al*., 2024), including in the context of the Notre Dame cathedral fire (Pasikowska-Schnass and Lim, 2023).

In the UK, there is an increasing interest in understanding the role of AI for the heritage sector. In a 2023 survey involving 154 members of the UK Heritage Pulse, respondents recognised the transformative potential of AI but raised concerns about challenges, including skills and funding shortages (Cantrill-Fenwick, 2023). 24% of respondents were aware of their organisation using AI, with "the highest proportion (41%) saying it was used to help with planning for the future (such as generating ideas), closely followed by marketing including generating content, editing, programming adverts and analysing or interpreting data" (Cantrill-Fenwick, 2023). Half of respondents had considered how AI may change how people interact with their organisation, including to research their organisation or area of heritage (30%) and to reproduce copyrighted materials (28%) (Cantrill-Fenwick, 2023).

It is unclear how the UK Heritage Pulse survey defined AI, and respondents may have focused more on commercial generative AI tools. The term "AI" is often perceived as a buzzword synonymous to generative AI tools such as OpenAI's Chat-GPT and Midjourney, which

---

[5] Stacey Patton, 'AI Meets Archives: The Future of Machine Learning in Cultural Heritage' (CLIR, October 21, 2024) https://www.clir.org/2024/10/ai-meets-archives-the-future-of-machine-learning-in-cultural-heritage/ [Accessed 18 February 2025]

[6] ERRIN (2024), 'Cultural Heritage & Tourism/ICT WG meeting: Unlocking the Power of AI for Cultural Heritage' https://errin.eu/events/cultural-heritage-tourismict-wg-meeting-unlocking-power-ai-cultural-heritage#:~:text=Through%20advanced%20AI%20algorithms%2C%20data,ensuring%20historical%20integrity%20is%20preserved [Accessed February 18, 2025]

dominate current policy discussions (as we will see in section 3), when in fact there is much more to AI beyond such commercial generative AI tools. Kretschmer *et al.* (2024, p. 119-121) distinguish pre-trained machine learning models (such as OpenAI's), and more bespoke researcher-trained models (which researchers may prefer for more accuracy and reduced bias), as well as the difference between Natural Language Understanding algorithms and Natural Language Generation algorithms (also known as Generative AI). Understanding how bespoke and non-commercial AI models are used as tools for research or collections management in heritage contexts is crucial, as they present different copyright law implications - particularly regarding "non-commercial research" - compared to commercial generative AI tools, as we will see in section 2.

In the UK, examples of bespoke uses of AI for heritage and collections management, use and research include: "making content, information and collections easier to find; generating new insights or knowledge from existing content, information and collections; and, supporting data collection, restoration and conservation work" (Pavis, 2023). The *FloraGuard* project, involving the Royal Botanic Gardens Kew, aimed to tackle illegal trade in endangered plants, and developed AI algorithms so that "the researchers could more efficiently search for and extract information relating to the illegal harvest and sale of endangered plants, from a range of cyber hotspots."[7] The *Living With Machines* project, involving the British Library, investigated the impact of technology on the lives of ordinary people during the Industrial Revolution, using machine learning to analyse data at scale.[8] The *Transforming Collections* project, involving the Tate, "combines critical art historical and museological research with participatory interactive machine learning design to surface suppressed histories, amplify marginalised voices and re-evaluate artists and artworks ignored or sidelined by dominant narratives."[9]

While AI presents valuable opportunities for heritage research, it also poses challenges, particularly in ensuring responsible use. Pavis (2023) identifies the following risks: bias,

---

[7] Led by the University of Southampton and partnered with UK Border Force and Royal Botanic Gardens, Kew. https://www.kew.org/read-and-watch/floraguard-plant-cyber-crime-illegal-wildlife-trade [Accessed April 30, 2024].

[8] Partnership between The Alan Turing Institute, the British Library, and the Universities of Cambridge, East Anglia, Exeter, Queen Mary University of London and King's College, London. https://livingwithmachines.ac.uk/about/ [Accessed April 30, 2024].

[9] Led by University of the Arts London and Tate. https://www.arts.ac.uk/ual-decolonising-arts-institute/projects/transforming-collections [Accessed April 30, 2024].

discrimination and misinformation; lack of transparency and traceability; undervalued contribution; human labour replacement; privacy, copyright and other rights infringement. This paper focuses on the copyright risks (and related issues such as transparency and bias) applicable to heritage research, a discussion which while emerging[10] remains underexplored.

Testing AI in heritage settings offers an important opportunity for assessing the benefits and risks of the technologies at stake. Museums, in particular, play a key role in critically engaging with AI and its impact, "by being open and accountable about what technologies they are using, and through public programs and contemporary collecting to develop visitor literacy around AI" (Murphy and Villaespesa, 2020). Heritage stakeholders (including researchers in such contexts) are thus ideally positioned to promote meaningful discussions on AI and the law, and to provide insights that can shape better policies and regulation, balancing the sector's AI use for cultural and societal benefits with respecting the core rights of creators. Therefore, it is important to understand what specific copyright risks and issues face heritage stakeholders (including researchers) when using AI.

2. COPYRIGHT CHALLENGES ARISING FROM AI USES IN THE CULTURAL HERITAGE SECTOR IN THE UK

2.1. The basics of UK copyright law, and how it applies to AI and heritage collections: copyright works, owners and rights

Training AI systems, including constructing corpora for machine learning, often involves using data protected by copyright such as texts, images, and videos (Iglesias Portela *et al*., 2019; Kretschmer *et al.*, 2024 p. 110).[11]

---

[10] See: A Hawkins, A.M. Sichani (2024). 'Data Matters': Report on the Towards a National Collection Discovery Projects focus group on data management, documentation, and archiving practices in digital cultural heritage projects. Towards a National Collection https://doi.org/10.5281/zenodo.14006955 and Anna-Maria Sichani (2024) A toolkit for Generative AI, data protection and intellectual property in digital cultural heritage. https://sas-dhrh.github.io/genai-cch-toolkit/ [Accessed February 18, 2025].

[11] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021) Para 18. https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

Ahnert *et al.* (2023) highlight copyright and contractual challenges in using digitised historical collections in the context of the *Living With Machines* project. They refer specifically to copyright legislation and digitisation funding in the UK (and resulting contractual agreements), creating a "mixed-rights" landscape resulting in access issues. They emphasise a "patchwork approach to digitisation", which can result in unrepresentative collections, potentially biasing AI research. Resolving these challenges requires systemic changes in funding priorities and national copyright policy, project-level solutions being insufficient (Ahnert *et al.*, 2023, p. 23). Given that the copyright status of materials constitutes one of the selection criteria for decisions on digitisation (Tolfo *et al.*, 2023, p. 31; Beelen *et al.*, 2023, p. 4), we believe copyright impacts on issues of bias, potentially affecting the quality of AI research. To explain this connection, we need to analyse basic copyright concepts.

Copyright law protects author's creations, categorised as "works", which in the UK are: "(a) original literary, dramatic, musical or artistic works, (b) sound recordings, films or broadcasts, and (c) the typographical arrangement of published editions."[12] Beyond the concept of "works", UK copyright law also protects performances[13] and *sui generis* database rights[14] (different from the copyright granted to authors of original databases). *Sui generis* database rights protect against unauthorised extraction and reutilisation of substantial amounts of a database that required substantial investment in obtaining, verifying or presenting the data (Kretschmer *et al.,* 2024, p. 124). As such, raw data or metadata (for example, dates, editions and ISBN) that may fall short of copyright protection as works could still be eligible to some degree of protection under the *sui generis* database right. Metadata such as reviews and summaries may also be eligible for copyright protection as works.

Copyright works have specific legal definitions, which do not necessarily resonate with non-specialised audiences. A "literary work" is not only a work of literature; it means any work "which is written, spoken or sung",[15] including books, journal articles, and other writings such as pamphlets, lectures,[16] and even computer programs and databases.[17] "Artistic works", some

---

[12] UK CDPA 1988, s. 1.
[13] UK CDPA 1988, s 180.
[14] The Copyright and Rights in Databases Regulations 1997, s. 13.
[15] UK CDPA 1988, s. 3.
[16] Berne Convention for the Protection of Literary and Artistic Works (adopted in 1886, entered into force 5 December 1887), art. 2(1).
[17] UK CDPA 1988, s. 3; WIPO Copyright Treaty (WCT) (adopted in December 20, 1996), art. 4 and 5.

protected irrespective of their artistic quality,[18] cover not only traditional visual artworks such as paintings, but also graphic works such as maps.[19] Translations, adaptations and musical arrangements are protected as works, and so are collections such as encyclopaedias and anthologies which, by reason of the selection and arrangement of their contents, constitute intellectual creations.[20] This shows the wide range of materials embraced by copyright that heritage organisations may have in their collections or engage with in their activities.

To qualify for copyright protection, literary, dramatic, musical and artistic works must meet the criterion of originality, i.e. reflecting the "author's own intellectual creation."[21] Copyright protection applies only to expressions, and not to ideas, procedures, operational methods or mathematical concepts as such,[22] nor to news and facts.[23] But even though data or information are not protected (Kretschmer *et al*, 2024), the form in which it has been expressed (as a news article and the typographical arrangement of the newspaper) will likely be.[24] Therefore, heritage collection items will very likely be considered "original" and expressive works protected by copyright.

The question of whether AI-generated works qualify for copyright protection or a similar right, and who should own it, remains debated (Hugenholtz and Quintais, 2021; Ramalho, 2017; Guadamuz, 2017). But this article has a different focus, i.e. whether copyright is infringed by the use of protected materials for heritage research utilising AI.[25]

---

[18] UK CDPA 1988, s. 4(1)(a).

[19] Berne Convention, art 2(1).

[20] Berne Convention, art 2(5).

[21] Infopaq International A/S v Danske Dagblades Forening (2012) Case no. C-5/08. In the UK, the originality test required that works reflected the creator's "skill, labour and judgement", but this has been surpassed by the EU *Infopaq* "authors' own intellectual creation" test: THJ Systems Limited & Anor v Daniel Sheridan & Anor [2023] EWCA Civ 1354.

[22] WIPO Copyright Treaty (WCT), art. 2; Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement), art 9(2).

[23] Berne Convention, art. 2(8); Walter v Steinkopff [1892] 3 Ch 489.

[24] See also CDPA, s 3(2): literary, dramatic or musical works must be recorded in writing or otherwise.

[25] The analysis of the copyright status of artistic works utilising AI tools (or even of the tools themselves) in heritage contexts may be the scope of future work, as these works are starting to integrate UK museum collections, such as the V&A recent acquisition of *MEMORY (Drawing Operations Unit Generation 2)* by Sougwen Chung: Kathrin Mitchell (V&A Blog, 14 December 2022) https://www.vam.ac.uk/blog/digital/the-algorithmic-gesture-sougwen-chungs-memory?srsltid=AfmBOoqdmjXhwgiHYop09besEAR-7i5nWI7NEsXnTsAqhoC0VM4YOM-i&doing_wp_cron=1728552593.7939109802246093750000. UK CDPA s 9(3) on computed-generated works: "the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken."

The first owner of copyright is usually the author, i.e. the person who creates the work.[26] Objects such as newspapers, books or CDs may thus embed multiple works with different copyright owners (Bently *et al.,* 2022, p. 62, 136-140). Ownership can also rest with employers,[27] or be assigned to third parties.[28] In many cases, information about the owner becomes lost, leading to the orphan works issue, which is significant for cultural heritage institutions (Korn, 2009). In a rights clearance simulation study, the British Library has estimated that over 40% of the potentially in-copyright works were orphan works (Stratton, 2011; Rosati, 2019).

Copyright protection generally lasts for the author's life plus 70 years,[29] after which the work enters the public domain. Until then, permissions are needed if the intended use falls under the exclusive rights of copyright owners.[30] Particularly relevant are the rights to copy the work[31] (reproduction right), to communicate a work to the public[32] and to make an adaptation.[33] Infringement occurs when someone, without permission, engages in such restricted acts in relation to a substantial part (qualitatively, rather than quantitatively: Rosati, 2019, p. 206) of the work.[34] Reproducing or communicating to the public even a small part of a work could infringe copyright, if that part represents the originality of the work, i.e. the "author's own intellectual creation" for literary, dramatic, musical or artistic works.[35] For entrepreneurial works (i.e. sound recordings, film recordings, broadcasts and typographical arrangements of published editions), copying any recognisable part may infringe copyright.[36]

These exclusive rights cover actions which are commonly part of AI development, training and usage, such as creating digital copies of works, further copies of already digitised works, sharing copies for verification, or showing in AI outputs parts of works. Text and data mining

---

[26] UK CDPA 1988, ss 9(1) and 11. As per s 9(2): that person is taken to be the producer (for sound recordings); the producer and the principal director (for films); the person making the broadcast (for broadcasts); and the publisher (for typographical arrangement of a published edition).
[27] UK CDPA, s. 11(2).
[28] UK CDPA, s. 90(1).
[29] UK CDPA 1988, s 12. Duration varies depending on the work.
[30] UK CDPA, s 16.
[31] UK CDPA, s 17.
[32] UK CDPA, s 20.
[33] UK CDPA, s 21.
[34] UK CDPA 1988, s 16(2) and (3).
[35] *Infopaq* Case C-5/08: taking even 11 consecutive words could constitute infringement.
[36] Pelham GmbH and Others v Ralf Hütter and Florian Schneider-Esleben (2019) Case C-476/17; ECLI:EU:C:2019:624. Contrast with earlier UK approach in England And Wales Cricket Board Ltd & Anor v Tixdaq Ltd & Anor [2016] EWHC 575 (Ch), which focused on whether the part reproduced reflected "investment".

(necessary for AI development and training) can implicate copyright if it involves making copies of works or breaching publishers' licences (Bently *et al.,* 2022, p. 260).[37] As such, these activities may require permission, unless an exception applies (which we analyse in subsection 2.2).

Copyright can thus hinder AI use in heritage research and management. Cultural heritage institutions house various types of copyright protected materials, which when digitised (an act which in principle requires copyright owner's permission) become important data for AI. Holding physical works does not grant heritage institutions copyright ownership, unless it was explicitly assigned to them (Heritage Digital, 2021). Furthermore, digitised collections (even of works in the public domain) may be subject to contractual restrictions, if digitisation was performed by a third party (Ahnert *et al.*, 2023, p. 27). Other restrictions include agreements with donors, heirs, loaning institutions, researcher agreements and website policies (Wallace, 2023).

The difficulties of clearing copyright for heritage collection digitisation is a topic of extensive academic and sector discussion. Copyright clearance can be costly and time consuming, to the point that heritage organisations may not be able to carry it out. Stratton (2011) explains that "rights clearance of works on an individual, item by item basis is unworkable in the context of mass digitisation". The lack of public funding results in digitisation of national assets to be undertaken by private companies that place limits on access (Ahnert *et al*, 2023, p. 5). Those unable to afford a licence to access materials for AI use may only be able to access cheaper and less reliable data, possibly resulting in biased AI results.[38] Margoni and Kretschmer (2022,

---

[37] Some scholars suggest a purposive interpretation of the reproduction right to exclude 'text and data mining' from the scope of copyright (Bently et al., 2022, p. 261). Margoni and Kretschmer (2022) argue that there should not be a need for a TDM exception for extracting informational value of protected works. Similarly, see Murray-Rust P. The Right to Read is the Right to Mine. Open Knowledge (2012) https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/ [Accessed October 7, 2024]. Concluding that the fact that TDM was regulated as an exception means that legislature sees these activities as falling under the remit of copyright, see Rosati (2019). A formalistic interpretation of the reproduction right, which currently prevails in landmark judicial interpretations such as *Infopaq*, would allow copyright owners to inhibit technical copies made through TDM: European Copyright Society, p. 5. Available from: https://europeancopyrightsociety.org/wp-content/uploads/2015/12/ecs-opinion-on-eu-copyright-reform-def.pdf [Accessed February 18, 2025].

[38] See Coalition for a Digital Economy (COADEC) response to UK IPO public consultation, noting prohibitive costs (time and money) for startups and scale-ups regarding difficulties in identifying ownership and multitude of rightsholders. See also CREATE response to public consultation. Available from: https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed February 19, 2025]. The costly and lengthy process of acquiring licenses is not a new issue, and was at the core of discussions on the implementation of the UK TDM exception introduced in 2014, see IPO's Impact Assessment Available from: https://www.legislation.gov.uk/ukia/2014/156/pdfs/ukia_20140156_en.pdf page 8.

p. 687) similarly highlight the risks involved in purchasing cheaper data or pre-trained models, which can result in biases and inaccuracies. Where there is uncertainty on the legality of scraping data, there is risk of copyright infringement, a situation which favours the development of "foundation" AI models "developed by the few large tech corporations which have access to the necessary data and can afford the uncertainties and costs of potential copyright litigation" (Kretchmer *et al*., 2024, p. 125; Margoni *et al*., 2022). This situation invites the consolidation of a "techno-economic oligopoly" and unsustainable (in legal, economic, social, cultural and environmental terms) practices of "data extractivism" or "data colonialism" (Kretchmer *et al*., 2024, p. 125; Couldry and Mejias, 2019).

Wallace (2021) highlights the risk-averse attitudes adopted in heritage management, digitisation, and online dissemination due to copyright complexities, given limited time, staff, and financial resources in the heritage sector. Additionally, in light of the heritage sector's role as custodians of culture and knowledge, the ethical, accuracy and reliability issues in AI-mediated research will discourage the use of lower-quality data by heritage professionals and researchers. In cases of legal uncertainty, projects will likely either be regulated under licensing terms for institutions that can afford them, or abandoned if licensing is unaffordable or impractical.

The biases, omissions and inaccuracies that may be generated by datasets produced or models trained on the basis of copyright permissions, particularly in heritage contexts, require further investigation. As highlighted in recent government consultation, "works being mined can be restricted by curatorial bias, only mining what is available under licence, rather than what would be most useful for the purposes of the AI."[39] Choosing the most affordable or easily accessible dataset may also not be in the best interest of the research question. The *Living With Machines* team addressed the challenges in timely obtaining digital data, noting this could have pushed them to work with more permissively available datasets, as pursuing their preferred dataset as determined by their research agenda (which they decided to do) required complex negotiations and legal expertise to navigate copyright and licensing issues; the team noted that the "current time-frames of this process are not compatible with publicly funded projects,

---

[39] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021) https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

which are by necessity time-limited in nature and assume a quick start from day one" (Ahnert *et al.*, 2023, p. 30-31). These kinds of resources, including of time and expertise, may not be available to other projects or institutions.

We believe that this encapsulates a key issue, in that the copyright status of the dataset will dictate what research can be made, limiting researchers' ability to select the most appropriate datasets to answer their questions. This imposes a bar to research engaging with more contemporary themes (such as digital humanities research on late nineteenth/early twentieth century onwards), as materials are more likely to be in copyright and thus require clearance (Ahnert *et al.*, 2023, p. 31). The quality of the input data is crucial to the machine learning process; researchers need to identify the necessary data aligning with the research purpose and might prefer to train their own models, as pre-trained embeddings rely on easily found text material leading to bias problems (Kretschmer *et al.*, 2024, p. 111-120).

It is not in the public interest that copyright law should affect the quality of research by imposing excessive barriers to data access. Copyright exceptions exist precisely to support activities in the public interest, such as access to culture, education, research, and freedom of expression (Rendas, 2018; Geiger and Izyumenko, 2020; Jacques, 2021; Bently *et al*., 2022). Our analysis will focus on examining these copyright exceptions to determine whether they are suitable for AI research in heritage contexts.

2.2. Copyright exceptions: text and data mining and beyond

Copyright exceptions often offer a clearer route for users requiring legal certainty, by allowing certain uses without the need for permissions.[40] Thus, exceptions assume a particularly relevant role in the risk-averse heritage sector (Hudson, 2020). Many copyright exceptions apply to activities of heritage professionals and researchers in those contexts.[41] Particularly relevant to AI research in the heritage sector is the UK exception for text and data analysis for non-commercial research[42] (hereafter "text and data mining" or "TDM").[43] This exception applies

---

[40] On the legal certainty of the EU TDM exception, see Rosati (2019) p. 214.
[41] CDPA 1988, ss.40A-43A. The UK revoked the EU orphan works exception following Brexit. The use of orphan works in the UK is regulated through a government licensing scheme: https://www.gov.uk/guidance/copyright-orphan-works.
[42] CDPA 1988, s. 29A.
[43] Using TDM interchangeably with text and data analysis when referring to the UK exception, see: Rosati (2019) p. 198; Bently, Sherman et al (2022) p. 260; UK IPO. Exceptions to copyright: Research. (2014)

to copyright works, though an equivalent exception is applicable to performances.[44] TDM is arguably also possible under the non-commercial research exception to the *sui generis* database right.[45]

The UK TDM exception, introduced in 2014, aimed to modernise UK copyright law for the digital age.[46] While it predates current AI discussions, admittedly the exception can encompass AI uses.[47] It was intended to support diverse research by being technologically neutral and not limited to academic papers or STEM fields.[48] Even though we believe this exception in principle supports AI heritage research, its application remains unclear and may be unsuitable in practice.[49] As noted in the *Living With Machines* project, the exception has proved difficult to use in innovative research involving diverse datasets, limiting its effectiveness in supporting national priority research in the intersection of technology and culture (Ahnert *et al.*, 2023, p. 28). The British Library response to the 2022 UK IPO Consultation noted that potential TDM projects have been abandoned on multiple occasions due to the inadequacy of the current s.29A

---

https://assets.publishing.service.gov.uk/media/5a7d678ee5274a02dcdf4502/Research.pdf [Accessed October 8, 2024]; and UK Intellectual Property Office. The Technical Review of Draft Legislation on Copyright Exceptions: Government Response. (HM Government, Intellectual Property Office, March 2014, DPS/IP B900-03/14). (2014) https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/response-copyright-techreview.pdf [Accessed October 8, 2024].

[44] CDPA, s. 1D in Schedule 2.

[45] Databases Regulations 1997, reg. 20. "The Government's view is that this existing exception [reg. 20] will permit the extraction of whole works if required for text and data mining through the provision for "*fair dealing with a substantial part.*" (page 13): https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/response-copyright-techreview.pdf. The most recent Government consultation (December 2024) flagged that the current UK TDM exception does not extend to databases: https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence [at para. 123].

[46] Hargreaves Review of Intellectual Property (2011) p. 9. https://researchbriefings.files.parliament.uk/documents/SN06430/SN06430.pdf [Accessed October 5, 2024] and the IPO's Impact Assessment. Exception for copying of works for use by text and data analysis BIS0312. 24 October (2012) https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/consult-ia-bis0312.pdf [Accessed October 5, 2024].

[47] Acknowledging text and data mining is necessary to develop and train 'artificial intelligence' algorithms: Bently *et al.*, (2022) p. 260. See also: Rosati (2019) p. 198; Strowel A. and Ducato R. (2021).

[48] UK IPO. The Technical Review of Draft Legislation on Copyright Exceptions: Government Response. (HM Government, Intellectual Property Office, March 2014, DPS/IP B900-03/14). (2014) https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/response-copyright-techreview.pdf [Accessed October 8, 2024].

[49] British Library response to the IPO Public Consultation (2022): https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024]. Response to the IPO Call for Views (2021), Prof Ruth Ahnert, PI of *Living With Machines,* as regards the TDM exception: "I would say it is not fit for purpose. … for cautious institutions with a high profile that ambiguity can be very limiting": https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views [Accessed February 23, 2025].

exception, requiring researchers to seek permissions from rights holders—an expensive and resource-intensive process, especially given the large volumes of content required for TDM.[50]

We advocate for addressing these issues through copyright exceptions, rather than licensing, as a more effective approach to promote public interest activities.[51] This approach would better address the resource constraints faced by the heritage sector and help mitigate the "curatorial bias" mentioned in section 2.1.

2.2.1. *Copies for computational text and data analysis: does it cover AI?*

The UK government defines TDM as the use of automated techniques to analyse text and data for patterns, trends, and insights, which typically involves the copying of works.[52] The introduction of a TDM exception for non-commercial research allows this copying without infringing copyright,[53] aiming to be technologically neutral and broadly applicable.[54]

Recent government consultations explored the applicability of the TDM exception to AI. While copyright owners argued the exception does not cover AI, others defended TDM as integral to AI development.[55] The government recognised TDM's role in AI systems used in research and by cultural heritage organisations.[56] TDM techniques are important in AI

---

[50] UK IPO. Artificial Intelligence and IP: Copyright and Patents Consultation Outcome. (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024].

[51] The JISC report cited in the IPO Impact Assessment document stated that: "the broader interests of equity may support the case for an exception to enable text mining so that society can maximise the potential returns from an asset in which society has made the lion's share of investment and taken the vast majority of the risk." https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/consult-ia-bis0312.pdf [Accessed October 5, 2024]; Diane McDonald and Ursula Kelly, The Value and Benefit of Text Mining. JISC (2012) https://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining [Accessed February 23, 2025].

[52] UK IPO. Exceptions to copyright: Research. (2014). https://assets.publishing.service.gov.uk/media/5a7d678ee5274a02dcdf4502/Research.pdf [Accessed October 8, 2024].

[53] UK IPO. The Technical Review of Draft Legislation on Copyright Exceptions: Government Response. (HM Government, Intellectual Property Office, March 2014, DPS/IP B900-03/14). (2014) https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/response-copyright-techreview.pdf [Accessed October 8, 2024] p 11.

[54] Ibid, p. 12.

[55] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021) https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

[56] UK IPO. Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation, Consultation outcome. (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation [Accessed October 8, 2024]

development, using the same algorithms to discover patterns in data (Rosati, 2019; Strowel and Ducato, 2021). Given the exception's technological neutrality, we believe that the concept of computational text and data analysis in UK law[57] comfortably encompasses AI development, training and use in heritage research contexts.

The types of copies allowed by this exception include those of a non-temporary nature. The "making of temporary copies" is specifically allowed in another exception,[58] and the TDM exception makes no reference to allowing only temporary copies. Permanent copies of training data can be "fundamental to the replicability of machine learning results" (Kretschmer *et al.,* 2024, p. 126). Having access to training data enables the detection of mistakes, omissions, or biases, ensuring greater transparency and accountability in decision-making (Margoni and Kretschmer, 2022, p. 688; Levendowski, 2018).

While UK law is silent on how long temporary copies can be retained, the TDM exception in the EU Digital Single Market (DSM) Directive[59] states that copies should be securely stored and retained for scientific research, including for the verification of results,[60] and that rightsholders, research organisations and cultural heritage institutions should be encouraged by member states to define commonly agreed best practices on this point.[61] Though the UK is not bound by the DSM Directive,[62] these requirements should be observed in the UK as best practice.

In the absence of statutory clarity, evidence is starting to emerge of arrangements imposing time limits for data to be kept in storage (of two years for example, see Ahnert *et al.*, 2023, p. 28). We suggest IPO Guidance to clarify reasonable time periods for copy retention, which should be determined in consultation with heritage and research stakeholders, on the basis of specific verification needs and storage feasibility. Engagement with such stakeholders could also help establish appropriate standards for copy retention. Ultimately, it may be necessary to update the text of the UK TDM exception to provide clarity on this issue.

---

[57] CDPA s 29A(1)(a).
[58] UK CDPA, s 28A. EU Directive 2001/29/EC (Infosoc Directive) art 5(1).
[59] EU Directive 2019/790 (DSM Directive)
[60] DSM Directive, art 3(2).
[61] DSM Directive, art 3(4).
[62] UK Parliament. Copyright: EU Action. (2020) https://questions-statements.parliament.uk/written-questions/detail/2020-01-16/4371 [Accessed October 6, 2024].

Furthermore, while the text of article 3, along with recitals 15 and 38 of the DSM Directive, represents progress in promoting transparency and accountability in algorithmic decision-making tools, there remains uncertainty regarding researchers' ability to grant access to stored copies for verification, as this may constitute an act of communication to the public, not exempted in the EU TDM exceptions (Margoni and Kretschmer, 2022, p. 697). We believe this issue should also be clarified in UK law, which similarly limits the TDM exception to the right to copy (or reproduction right). If we understand that copies can be retained for research verification, the exception should be interpreted to permit this. One way this could be achieved is by interpreting the concept of "public" (in "communication to the public") as not including individuals seeking access for research verification under specific circumstances. Further research and policy work are required on this, involving extensive engagement with heritage and research stakeholders in the UK.

The UK TDM exception only allows the making of copies and not the dissemination (i.e. "communication to the public") of copyright works. This limitation, noted in the Technical Review of Draft Legislation on Copyright Exceptions (2014) means that TDM cannot result in making full copyright works publicly available. Margoni and Kretschmer (2022) argue that confining the exception to reproduction is too restrictive. That said, other exceptions such as the quotation exception,[63] may permit the communication to the public of excerpts of a work. Government guidance confirms that if parts of a work need to be quoted in TDM research outputs, the quotation exception can apply, so long as copyright laws are followed.[64]

A last point concerns the types of "copies" the legislation refers to, specifically whether it covers the digitisation of physical materials, or only applies to already digitised works. This is relevant to heritage collections, which contain vast amounts of undigitised materials of value to TDM research. The UK exception does not specify the format of the work to be copied, arguably allowing for the possibility of making digital copies of physical works as long as there is "lawful access" to the work. We therefore find that, ultimately, the analysis of whether the law allows digitisation of physical works for TDM conflates with the concept of lawful access, which we analyse below.

---

[63] S 30(1ZA) UK CDPA.
[64] UK IPO. Exceptions to copyright: Research. (2014), p. 9. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf [Accessed October 6, 2024]

2.2.2. "Lawful access" and the effectiveness of the exception

The concept of lawful access within the TDM exception requires clarification. Key issues include whether physical access to works, rather than digital access, qualifies for TDM use, and how contractual terms may override the exception. We will also look at how Technological Protection Measures (TPMs) may render the exception ineffective. Lastly, we will analyse the prohibition of transferring copies as a defining aspect of lawful access, and how unsuitable this is to contemporary forms of collaborative digital heritage research.

(a) a lawful "digital" access?

The meaning of lawful access in the TDM exception is unclear, particularly regarding whether it applies solely to digital access or includes access to physical works, thereby allowing digitisation. We believe this point is especially relevant in heritage research, where collections often include physical works, in addition to the digital repositories typically used in academic research. While current scholarship on TDM and copyright focus on electronic data, they rarely address the digitisation of physical works or the specific needs of heritage research.

The UK exception does not define lawful access. As such, it does not qualify access as only *digital* access. This suggests that the creation of copies through the digitisation of a physical object, to which the copier has lawful access through having lawfully acquired it, should be allowed by the exception. In the case of a heritage organisation, we believe lawful access would include having legal ownership of objects through lawful acquisition (including bequests, field collection, gifts, purchases, exchanges and treasure).[65] This interpretation aligns with the outcome of the *Google Books* case in the US, where digitisation of physical books owned by libraries was considered fair use for TDM purposes.[66]

---

[65] The Collections Trust *Spectrum* standard of museum collections management in the UK (also used worldwide) explain the procedures for acquisition and accession: https://collectionstrust.org.uk/spectrum/procedures/acquisition-and-accessioning-spectrum-5-0/ and https://collectionstrust.org.uk/resource/acquisition-and-accessioning-suggested-procedure/ [Accessed February 23, 2025].

[66] Authors Guild v Google, Inc, No.13–4829 (2d Cir. 2015), affirming Authors Guild v Google, Inc, 954F. Supp. 2d 282 (2013). See also Rosati (2019).

US fair use is a far more flexible and adaptable exception to copyright than those in the UK and EU, where the tradition of a strict interpretation of exceptions is followed,[67] as long as the effectiveness of the exception is not compromised and its purpose is achieved.[68] We therefore once again delve into the UK law makers' intentions and justifications for the introduction of the TDM exception, to assess whether creating digital copies of physical works for TDM purposes would fall under the exception as "lawful access".

The UK Government's 2012 response to the public consultation on the introduction of new exceptions, including TDM, left open the meaning of a prior "right to access", wording it as "under a licence or otherwise" and providing as examples "subscription to a scientific journal or having copies of papers published under a Creative Commons licence".[69] In our view, this allows *any* kind of prior right to access, digital and physical alike. What appears important is that the exception should not undermine the publishers "control over IT systems or commercial exploitation", and how unlikely it was that TDM copying would substitute the works.[70]

The IPO Impact Assessment stated that: "data analytics methods extract data from existing electronic information."[71] It further adds that "Copyright is not intended to prevent use of facts for research, and this exception is intended to remove the block on reuse of materials for research using these tools."[72] The IPO also noted the exception would apply "in cases where access to articles and / or data has already been gained (eg by subscription)."[73] It appears that the IPO's position focused on the public benefit in "more and higher quality research", and a lowering of costs and simplification of procedures for researchers, while also offering "incentives" and "security" for publishers, and a protection against undermining their primary market for access to works.[74]

---

[67]*Infopaq* (Case C-5/08).

[68] *FA Premier League Ltd* (C-403/08 & 429/08) [at 162-163].

[69] HM Government. Modernising Copyright: A Modern, Robust and Flexible Framework. Government Response to Consultation on Copyright Exceptions and Clarifying Copyright Law. (2012) p. 5 and 37 https://data.parliament.uk/DepositedPapers/Files/DEP2012-1936/ModernisingCopyrightGovernmentresponse.pdf [Accessed 6 October, 2024]

[70] Ibid.

[71] UK IPO. Exception for copying of works for use by text and data analytics. Impact Assessment. IA No: BIS0312. (2012) https://www.legislation.gov.uk/ukia/2014/156/pdfs/ukia_20140156_en.pdf [Accessed September 30, 2024].

[72] Ibid.

[73] Ibid.

[74] Ibid.

UK IPO Guidance explains that the new TDM exception allows "researchers to make copies of any copyright material for the purpose of computational analysis if they already have the right to read the work (that is, work that they have "lawful access" to)."[75] The Guidance highlights that researchers would "still have to buy subscriptions to access material; this could be from many sources including academic publishers."[76] The Guidance FAQ finally defines lawful access as covering cases: "where researchers have the legal right to access a copyright work to read it; examples could include paying for a subscription to a journal or database or material published under open licences including Creative Commons and Open Government Licences."[77]

While equating the new TDM exception concept of lawful access with cases where researchers would have already had the right to read the work including through purchasing subscriptions, the guidance does not specify that this only applies to the right to read the work digitally or to digital subscriptions, thus possibly allowing physical access as lawful access. We note that journal subscriptions can be print or electronic.[78]

Whether lawful access should mean access to digital platforms or repositories, or if it includes physical access to works is neither sufficiently discussed in academic literature, nor expressly addressed in legislation, government or sector guidance. We believe that the wording used in policy papers and IPO guidance is open enough to allow an interpretation that lawful access can include physical access. We should also note Recitals 10 and 14 of the EU DSM Directive, openly defining "lawful access", adding "other lawful means" of access.

If the legislator had intended to delimitate the mode of access to digital access, it would have done so expressly in the statutory text. We believe that the lawful acquisition of the material (be it physical or digital) by the heritage organisation, combined with the specific focus of the exception on text and data analysis for non-commercial research, offer sufficient safeguards to the commercial interests of copyright holders. This is in line with the "three-step test" in international copyright law, which dictates that exceptions should be confined to "certain

---

[75] UK IPO. Exceptions to copyright: Research. (2014), p. 6. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf [Accessed October 6, 2024].
[76] Ibid.
[77] Ibid, p. 7.
[78] See for example the Elsevier print and electronic subscription modes: https://www.elsevier.com/en-gb/about/policies-and-standards/pricing/journals [Accessed 23 February, 2025].

special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder."[79]

We therefore argue that "lawful access" should include either physical or digital access, thus allowing for the digitisation of physical materials for non-commercial TDM research. This solution meets the exception's objective, while allowing space for TDM to adapt in time and embrace new uses. Allowing the digitisation of physical works may help remedy issues of bias and gaps in digital collections,[80] thereby improving the quality of AI research and systems development.

(b)   *Lawful access, no contractual override of the exception, no transfer of copies and TPMs*

The TDM exception in UK law contains an important caveat, denying any contractual override of the exception.[81] However, in practice, the uncertainty about what "lawful access" means and the scope to regulate it through licences can undermine the exception's effectiveness. Lawful access is a paradoxical requirement, which can subvert the innovative aims of the TDM exception (Kretschmer *et al.*, 2024, p. 131) and represents a restriction on the enjoyment of the exception if interpreted to always depend on the terms of a contract or licence (Synodinou, 2019, p. 27). This makes the exception subject to private ordering (Geiger *et al.*, 2019; European Copyright Society, 2017; Margoni and Kretschmer, 2022, p. 697), as "the exception can effectively be denied to certain users by a right holder who refuses to grant 'lawful access' to works or who grants such access on a conditional basis only" (European Copyright Society, 2017, p. 4). It can also reflect on access licensing pricing, by allowing publishers to price TDM into subscriptions fees, which many organisations will not be able to acquire (European Copyright Society, 2017, p. 4). Geiger *et al.* (2019) and Bottis *et al.* (2019) argue that lawful access for TDM increases the cost of research, potentially pricing out underfunded institutions, exacerbating existing inequities in scientific and technological development. Also, "lawful access" can "severely impair other fundamental rights such as the freedom of information and

---

[79] TRIPS Agreement, art. 13; Berne Convention, art. 9(2). The WTO interpreted the test as requiring exceptions to be clearly defined and narrow in scope and reach (Panel Rep. of 15 June 2000, WT/DS160/R, p. 34). Geiger *et. al.* (2010) promoted a declaration, with which we agree, on the need for a broader, more balanced application of the test, requiring a comprehensive overall assessment, balancing interests of rights holders and general public. See also Westenberger (2017, p. 296, 301-302, 305-306) and Hudson (2020, p. 14-19).

[80] E.g. the need to digitise Mitchell's NPD (Ahnert *et al.*, 2023, p. 33-34).

[81] S. 29A(5) CDPA.

to inform the public about specific undisclosed but publicly relevant issues" (Margoni and Kretschmer, 2022, p. 697. See also Dusollier, 2020, p. 987).

We believe uncertainties around the concept of "lawful access" are problematic within risk-averse heritage research contexts. The European Copyright Society (2017, p. 4) explains legal uncertainty prevents risk-avoiding beneficiaries from relying on exceptions, noting also that the benefit of exceptions, particularly based on fundamental rights or public interests such as research exceptions, should not be dependent on market decisions of copyright owners.

Furthermore, the TDM exception prohibits transferring copies of works to any other person unless authorised by the copyright owner.[82] This restriction is unclear[83] and misaligned with the collaborative nature of modern research in heritage contexts, where organisations holding relevant data often partner with those possessing the computational resources and expertise needed for AI projects. Research practices have evolved significantly since the exception's introduction in 2014, and this requirement has not kept pace, potentially forcing research collaborations to rely on licensing agreements for data transfer or storage with repositories. The limitations of "lawful access" became evident in the *Living With Machines* project, where a key challenge "was the transfer of data between spaces, from the data owner (BL) to the owner of the infrastructure (The Alan Turing Institute)", and researchers had to negotiate a bespoke agreement with a commercial partner (FindMyPast) to access digitised data from the British Library holdings, even though the British Library were project partners (Ahnert *et al.*, 2023, p. 28). Even materials that are out of copyright may be subject to contractual restrictions if they have been digitised by a third party, as seen in the case of the British Newspaper Archive, which was digitised by FindMyPast (Ahnert *et al.*, 2023, p. 27).

Kretschmer *et al.* (2024, p. 131-132) argue that the lack of clarity around the conditions for copying in machine learning contexts is likely to affect scientific research, highlighting that "lawful access" terms will dictate what research is possible and at what cost, meaning that research which should benefit from the copyright exception would instead be governed by licensing agreements. In such cases, rightsholders could threaten to withdraw access from

---

[82] CDPA, s 29(A)(2)(a).
[83] See BBC's response asking for "clarification as to the application of s29A(2)(a) … where the entity commissioning the research (or indeed a third party) collects the data": UK IPO. Artificial Intelligence and IP: Copyright and Patents Consultation Outcome. (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024].

institutions. To illustrate the above, Kretschmer *et al.* provide an example involving research and heritage organisations with lawful access to broadcasting or newspaper digital archives: rights holders could choose to license these materials to AI companies for machine learning, potentially threatening to revoke access to archives in settings where they are utilised for research serving public interests. Kretchmer *et al.* (2024, p. 127) also argue that the power asymmetry of certain markets, combined with the techno-legal uncertainty that they discuss, may operate a *de facto* circumvention of the mandatory nature of the TDM exception in art 3 EU DSM Directive through access condition practices using Application Programming Interfaces (APIs).

Even though the exception cannot be contracted out, "licences may still impose conditions of access to the licensor's computer system, for example to maintain security or stability" (Bently *et al.*, 2022). These access conditions may also be achieved by the implementation of TPMs.

Lawful access to a database does not permit the circumvention of TPMs put in place by the rightsholder to safeguard it. However, rightsholders' use of TPMs must still adhere to the principle of proportionality (Recital 16 of the DSM Directive). Measures applied by rightsholders cannot go beyond what is necessary for the security and integrity of the networks and databases (art 3(3) DSM Directive). Despite this, "security controls, if too invasive, might deter or encumber legitimate TDM activities by researchers" (Dusollier, 2020, p. 296). In the UK, a remedy is available in case of TPMs "abuse", where technological measures prevent acts that are permitted (such as copyright exceptions).[84] The solution in the UK, however, is that a notice of complaint should be issued to the Secretary of State, and we have no knowledge of this remedy having been used.

Ultimately, these lawful access conditions (technological or contractual) should not undermine the effectiveness of the exception by creating excessive barriers or legal uncertainty. For this reason, the IPO Guidance on the TDM exception states that publishers and content providers can implement reasonable measures to maintain network security or stability, but these should not prevent or unduly restrict researchers from making necessary copies for text and data

---

[84] UK CDPA, s. 296ZE.

mining. Any contract terms that prevent researchers from making copies of works to which they have lawful access for TDM purposes will be unenforceable.[85]

We believe that "lawful access" conditions, including API and TPM practices, need further investigation to assess their impact on heritage sector research under the UK TDM exception. Additionally, we suggest revising the prohibition on transferring copies to expressly insulate research collaborations and practices (such as funder requirements) that require such transfers, which could also help mitigate barriers for smaller institutions and researchers with limited resources.

(c)   Does "lawful access" allow web scraping?

A key point of uncertainty in UK law is whether web scraping is permitted. While Recital 14 of the DSM Directive clearly states that lawful access should cover freely available online content, UK law lacks clarity. Rosati (2019) argues that activities like web mining, which use data mining techniques to extract knowledge from web data, may not require authorisation from copyright holders. IPO Guidance states that examples of lawful access "could include" materials accessed via subscriptions or open licenses,[86] which could suggest (particularly for the more risk averse) that web scraping may not be allowed unless the content is licensed as open access or depending on the terms of use of the website. IPO Guidance should be clearer on this point. We believe consultation with heritage sector stakeholders and researchers is needed to understand web mining practices and challenges, and how copyright law can better support heritage-related web research.

2.2.3.   "Non-commercial research": what of public-private collaborations and heritage management?

The TDM exception's restriction to non-commercial research raises two key issues for AI usage in the heritage sector. First, it reflects the uncertainty on what "non-commercial" means,[87] particularly given current policy calls for heritage organisations to commercialise

---

[85]   UK   IPO.   Exceptions   to   copyright: Research.   (2014) https://assets.publishing.service.gov.uk/media/5a7d678ee5274a02dcdf4502/Research.pdf [Accessed October 8, 2024]
[86] IPO Guidance on Exceptions for Research, p. 7.
[87] Boundaries between commercial and research are increasingly vague: Ariadna Matas, AI 'opt-outs': should cultural heritage institutions (dis)allow the mining of cultural heritage data? https://pro.europeana.eu/post/ai-opt-

their digital assets to generate income (as per the Mendoza Review), and also in the context of the public-private collaborations, for example in cases of heritage digitisation undertaken by commercial partners. We will consider whether "non-commercial" should be replaced with "non-profit". Second, we will explore the meaning of "research" to assess whether certain heritage management practices fall within this concept, and whether a new exception is needed.

(a)   "Non-commercial": a bar to public-private partnerships and heritage organisations income streams?

Heritage institutions are increasingly expected to explore new income streams through digital opportunities and may also collaborate with other partners, including commercial ones, in TDM/AI projects. It is not clear if the UK TDM exception for "non-commercial" research would apply to these projects.

The EU DSM Directive explains in Recital 28 in relation to the preservation exception that heritage institutions do not necessarily have the technical means or expertise to undertake digital preservation, and might need the assistance of other cultural institutions or other third parties for that purpose, and as such cultural heritage institutions should be allowed to rely on such third parties acting on their behalf and under their responsibility. We believe that a similar situation can be seen in relation to AI research.

It has been argued that the EU exception in art 3 of the DSM Directive includes both commercial and non-commercial research (Rosati, 2019, p. 212). Provided that the research is done by the parties allowed in the exception, i.e. "research organisations and cultural heritage institutions" (and the commercial partner has no decisive influence and control over the research organisation)[88], and it is for the purposes of "scientific research", there is no additional requirement in the EU exception that this research should be non-commercial.

UK IPO Guidance explains that the non-commercial nature of the research does not prohibit the publication of research outputs in commercial publications, but cautions researchers on the

---

outs-should-cultural-heritage-institutions-dis-allow-the-mining-of-cultural-heritage-data [Accessed February 19, 2025].

[88] Recital 12 and art. 2(1) DSM Directive.

need to carefully assess the original purpose of the research.[89] This opens the door to the idea that "non-commercial" research may include certain commercial aspects, although what these aspects might be remain unclear, posing challenges for researchers, in public-private partnerships (PPPs), where the boundary between commercial and non-commercial becomes blurred.[90]

The academic literature also acknowledges this uncertainty. Aplin and Davis (2021) address the complexities surrounding the definition of "non-commercial" research, questioning whether research must be entirely free from commercial intent to qualify. Brown *et al.* (2022) discuss the evolution of the non-commercial research exception under UK law, noting that the restriction to non-commercial purposes has been a point of contention. This ambiguity has affected both academic and professional research environments. Brown *et al.* suggest, and we agree, that the purpose of the research is to be assessed at the time it is conducted, and that it should be sufficient if there is 'a' non-commercial purpose, while noting the extensive ambiguity in the distinction between commercial and non-commercial research.

One case that precedes the UK TDM exception addresses the issue under the fair dealing exception for non-commercial research (s 29(1) CDPA). In *Controller HMSO and Ordnance Survey v Green Amps* [2007] EWHC 2755 (Ch) the court determined that research by a commercial entity could not be considered non-commercial, even if it lacked an initial commercial purpose. If courts today had to interpret the UK TDM exception for non-commercial research, they would likely apply this precedent to delineate that TDM research conducted by commercial entities should not benefit from the exception. However, it remains unclear whether non-commercial entities, such as universities and heritage organisations, can collaborate with commercial entities on specific aspects of the research.

We believe UK law should embrace the reality of public-private collaborations in TDM/AI research, ensuring that important collaborative efforts are not excluded from the exception. The impact on heritage research and potential solutions for legal interpretation—and, if necessary,

---

[89] UK IPO. Exceptions to copyright: Research. (2014), p. 10.
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf [Accessed October 6, 2024]
[90] See ABPI response to IPO public consultation on this point https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024].

changes to the law—warrant further empirical research and focused policy consultation, which we will explore in more detail in section 3.

The difficulties in defining "non-commercial" research are longstanding, as seen in earlier contexts, such as s 29(1) CDPA, and the orphan works policies as per government response in 2012.[91] In light of these difficulties, we believe that further research and policy work should consider whether instead of "non-commercial research", the exception should be framed to apply to research led by non-profit organisations, in the same manner that other heritage sector exceptions are framed in UK law, such as s. 40(A)(2) CDPA which allows the "lending of copies" by "public libraries" and by those libraries and archives "not conducted for profit".[92] This would arguably constitute a more objective standard than having to determine whether the research is "non-commercial".

Brexit provides flexibility to pursue a new exception for TDM that would not need to conform with the limits of EU law. The UK would however need to adhere to international conventions that set limits and minimum standards to national copyright legislation. Importantly in this case is the already mentioned three-step test, and our preferred more balanced interpretation of the test (under 2.2.2.a). It is imperative to understand, in a heritage context, what are such *special* cases that would require an exception, and what is a *normal* exploitation of works in those contexts. Heritage sector stakeholders (including researchers in those contexts) can also elucidate on the *reasonability* of the prejudices to rightsholders *legitimate* interests. We believe reasonability is a relative concept, and in the context of an exception should involve an assessment of harms (for example, limitation on income streams for rightsholders) *versus* benefits (for example, better research, preservation of heritage and access to culture).

The UK has consulted on the need for a broader exception,[93] with policy work reflecting this. A key issue has been the non-commercial aspect of the current exception. Among the responses received, AIPPI UK highlighted that s. 29A presents difficulties for non-commercial

---

[91] HM Government, Government Policy Statement: Consultation on Modernising Copyright (2012) p. 8 https://assets.publishing.service.gov.uk/media/5a7e07e3ed915d74e6223a7b/copyright-consultation-government-policy-statement.pdf [Accessed October 6, 2024].
[92] See also CDPA, ss. 42(4), 42(A)(1), and 43(A)4.
[93] UK IPO. Artificial Intelligence and IP: Copyright and Patents Consultation Outcome. (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024].

organisations collaborating with commercial entities on AI development.[94] Similarly, the BBC raised concerns about the exception's limitations when non-commercial entities outsource TDM to commercial third parties, particularly around lawful access and the lack of clarity on whose "purpose" should be relevant to define whether the research is commercial or not.[95] The European Alliance for Research Excellence also noted that the UK exception is too narrow for today's research landscape, especially in public-private partnerships, which discourages TDM activities due to copyright concerns.[96]

Gathering further evidence from heritage stakeholders is crucial for policymakers to create accurate guidance or legislative solutions for the TDM exception. It is important to address one final issue: the concept of "research" and whether it applies to certain heritage management practices.

(b)   "Research": what of heritage management?

A last point should be made as regards the copyright exceptions landscape to support AI uses in heritage contexts, which relates to the framing of the exception for research only, and as such it could exclude other types of AI usages for heritage management such as for cataloguing, preservation and reconstruction efforts.

It is important to remember that the exception was not intended to apply only for "scientific" research, with a discussion on the inclusion of the qualifier "scientific" as having a possible consequence of leading to the incorrect conclusion that the exception would only apply to academic papers or STEM research.[97] We believe that a broader interpretation of "research" could encompass certain heritage management activities aimed at internal organisational and collections management improvements.

Given the importance of heritage cataloguing data for research quality and accuracy, it is crucial that the TDM exception supports heritage AI projects aimed at improving this data. For instance, the *Transforming Collections* project, part of the UKRI/AHRC *Towards a National*

---

[94] Ibid.
[95] Ibid.
[96] Ibid.
[97] UK IPO. The Technical Review of Draft Legislation on Copyright Exceptions: Government Response. (HM Government, Intellectual Property Office, March 2014, DPS/IP B900-03/14). (2014) p.12 https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/response-copyright-techreview.pdf [Accessed October 8, 2024] p. 12.

Collection programme, is a clear research initiative. However, it also has the potential to provide AI tools useful to the heritage sector beyond the project's life. We believe that future uses of such tools in heritage organisations should be covered by the TDM exception, enabling internal research and enhancing collections data for future AI-driven research.

The future of AI research based on collections data looks promising, especially with national initiatives to unify collections in the UK. In addition to the *Towards a National Collection* programme, a key development is the launch of the Museum Data Service (MDS). This collaboration between Art UK, Collections Trust, and the University of Leicester will unify over 100 million museum records, offering the most comprehensive dataset of the nation's museum holdings. The MDS will be a vital resource for researchers, educators, curators, and content creators. As noted by the Minister of State for Science, Innovation, and Culture, this initiative enhances museums' digital capabilities, creating new opportunities for research, collaboration, and preservation.[98]

Recent case law in Germany, such as *Robert Kneschke v LAION gemeinnütziger e.V.* (case No. 310 O 227/23),[99] supports the idea that creating datasets for AI training through web scraping publicly available images can be considered research, as it contributes to future knowledge generation.[100] How similar reasoning would be considered in UK courts is uncertain. But we believe this case supports the understanding that certain heritage management TDM practices that contribute to future AI research (e.g. by improving cataloguing data) may also fall under the research exception. The question of whether the heritage sector needs a new exception to cover further AI-based heritage management activities requires further research and policy development.

While academic perspectives are valuable, they often remain untested in court and thus may not provide the legal certainty that stakeholders need to confidently rely on an exception. As of the writing of this paper, no case law has considered the (in our view unclear) s. 29A TDM

---

[98] Knowledge Integration. Introducing the Museum Data Service. https://www.k-int.com/introducing-the-museum-data-service/ [Accessed October 7, 2024].

[99] EUIPO, https://www.euipo.europa.eu/en/law/recent-case-law/germany-hamburg-district-court-310-o-22723-laion-v-robert-kneschke [Accessed February 19. 2025].

[100] Simon Hembt *et al*. Long-awaited German judgment by the District Court of Hamburg (Kneschke v. LAION) on the text and data mining exception(s) https://www.twobirds.com/en/insights/2024/germany/long-awaited-german-judgment-by-the-district-court-of-hamburg-kneschke-v-laion [Accessed February 19. 2025].

exception in the UK. It is thus important to delve into current law reform debates, including on this provision.

3. CURRENT COPYRIGHT AND AI POLICY AND REGULATION EFFORTS IN THE UK: SCOPE FOR FURTHER HERITAGE SECTOR PARTICIPATION

The UK government's ambition to lead in AI innovation and research,[101] has prompted significant policy work and stakeholder engagement, including how copyright law can support this goal. Given that many AI uses in heritage contexts are research-driven (Section 1), the views of stakeholders in the heritage sector could offer important insights into how copyright law supports or hinders the sector's use of AI, and what regulatory improvements are needed. This section will analyse the policy work undertaken in the UK, and assess the extent to which the heritage sector has been included in discussions based on publicly available data (including scholarly papers, policy documents, public consultation responses and official reports).

The UKIPO led a significant policy effort in 2020 by publishing a call for views on artificial intelligence and intellectual property.[102] The 2021 outcome report noted that 92 responses were received,[103] coming from various stakeholders, including copyright owners, creative and technology industries, licensing bodies, legal representatives, and academics.[104] Notably absent from this list were cultural heritage sector stakeholders, perhaps due to the small number of heritage respondents[105] or the conflation of the heritage sector with creative industries (which we do not believe is appropriate). The task of classifying respondents into categories is complex, as there may be overlaps - for example, although heritage organisations are often

---

[101] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021). Para 23 and 25 https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

[102] Ibid.

[103] Ibid. Para 6. We have found 88 files, however 3 respondents submitted 2 responses each, totalling 85 individual respondents: https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views [Accessed February 23, 2025].

[104] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021). Para 8. https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

[105] Archives and Records Association, European Alliance for Research Excellence, LACA – Libraries and Archives Copyright Alliance, and Ruth Ahnert (PI of *Living with Machines*). We should also note related organisations BAPLA and Creative Commons, noting however these do not represent only heritage stakeholders.

classed as "users", in some cases they may also fall under a rightsholders category.[106] It would have been important to see a deliberate mention to the heritage sector category both in the call for views and the government response.

The government response did not mention heritage stakeholders, or heritage uses of AI, except the more general remarks on benefits for researchers and creators, the risks for human creators, and the need to ensure that measures implemented will encourage AI for the public good while protecting intellectual property rights; and that in doing so the UKIPO will collaborate with "experts from business, technology and research" and "developers and users of AI and owners and users of intellectual property".[107]

One of the next steps outlined was to review how copyright owners license their works for AI use and explore ways to improve licensing or copyright exceptions to support innovation and research.[108] Indeed, conflicting responses to the call for views were given by copyright owners and users on the matter of use of copyright material for training AI. The preferred approach for most copyright owners was a voluntary licensing model, arguing that it would better balance remuneration with AI access. Many felt that current copyright exceptions do not apply to machine learning processes and that a licensing model would offer greater certainty.[109] Many copyright owners expressed concerns about moving towards an exception that would allow commercial TDM.[110]

On the other hand, users of copyright materials, including "technology firms, entrepreneurs and researchers", noted the disadvantages of relying on licences, including the high costs, which may only be affordable for "established or large businesses", and the curatorial bias that may be generated by only mining content "available under licence, rather than what would be most useful for the purposes of the AI".[111] The focus on "businesses" in the discussions fails

---

[106] See the BAPLA response, where heritage organisations are owners of photographs rights https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024].

[107] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021). https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

[108] Ibid. (Next steps - action: copyright. Para 5).

[109] Ibid.

[110]Ibid.

[111] UK IPO. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome. (2021). https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-

to consider public heritage organisations, such as museums and libraries, which face similar financial challenges and would struggle to afford licensing fees for AI-related projects.

Following the initial call for views, the UK IPO issued a public consultation on how AI should be addressed in the patent and copyright systems, receiving 88 written submissions from sectors such as the creative industries, technology, pharmaceuticals, the third sector, academia, and legal and IP professions,[112] but few from heritage stakeholders.[113] Despite this, the Government's response acknowledged that TDM is used for training AI systems and has applications in research, journalism, business analytics, and by cultural heritage organisations, and proposed expanding the scope of the TDM exception to permit TDM for any purpose (including commercial), while safeguarding rightsholders to protect their content, including a requirement for lawful access.[114] As seen in section 2, the requirement of "lawful access" is in itself extremely unclear, and it was not clear how the government would address such ambiguities.

Concerns raised by the creative industries and parliamentarians led to the abandonment of this reform. In 2023, the UK Minister for Science, Research, and Innovation confirmed that plans to broaden TDM exceptions had been shelved. The House of Lords Communications and Digital Committee's 2023 report on the creative industries recommended pausing the proposed changes and conducting an impact assessment on the potential effects on the creative sector, with industry groups arguing that weakening copyright protection could harm creators by reducing incentives for future investment in their work. The Committee noted that while AI development is important, it should not be pursued "at all costs,"[115] and any changes to the TDM regime must be balanced against the interests of the creative industries. Though the report briefly mentioned museums and galleries digitising collections and referenced research council

---

call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

[112] UK IPO. Artificial Intelligence and IP: Copyright and Patents Consultation Outcome. (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed: May 1, 2024].

[113] Archives and Records Association, British Library, European Alliance for Research Excellence, LACA, National Library of Scotland and Wellcome Trust. We should also note related organisations such as BAPLA, noting however it does not only represent heritage stakeholders.

[114] UK IPO. Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation [Accessed October 8, 2024].

[115] House of Lords Communications and Digital Committee. Chapter 2: A Digital Future. (2023) https://publications.parliament.uk/pa/ld5803/ldselect/ldcomm/125/12505.htm#_idTextAnchor019 [Accessed February 3, 2024]

programmes supporting heritage digitisation, heritage sector concerns were largely absent from the decision to halt the new TDM exception.[116]

The UK Government's AI White Paper outlined its plan to balance IP protection with AI development, by following Sir Patrick Vallance's recommendations.[117] Vallance recommended that "Government should announce a clear policy position on the relationship between intellectual property law and generative AI to provide confidence to innovators and investors" stating "an urgent need to prioritise practical solutions to the barriers faced by AI firms in accessing copyright and database materials."[118] Vallance added that "government should work with the AI and creative industries to develop ways to enable TDM for any purpose, and to include the use of publicly available content including that covered by intellectual property as an input to TDM (including databases)."[119] We find that the "AI and creative industries" framing does not sufficiently contemplate the heritage sector, and policy language should be revisited to engage heritage stakeholders more expressly. This would allow the understanding of what are the relevant AI technologies at stake (rather than only focussing on generative AI), barriers faced by this specific sector (rather than only "AI firms"), and how a new TDM exception could contemplate the specific uses by such stakeholders.

In response to the Vallance recommendation, the UK IPO instructed the drafting of a code of practice to help AI firms access copyrighted materials while protecting creators' rights.[120] However, the working group formed to create this code lacked diverse participation and the heritage sector was underrepresented based on the publicly available members list (the British Library was the only heritage organisation listed, against a high number of rightsholders and technology stakeholders).[121] This limited representation of the heritage sector neglects critical

---

[116] Ibid.

[117] Department for Science, Innovation & Technology. Policy paper: A pro-innovation approach to AI regulation (2023). Para. 35. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [Accessed February 23, 2025].

[118] HM Government. Pro-innovation Regulation of Technologies Review: Digital Technologies. Report presented by Sir Patrick Vallance. Recommendation 2. (2023) https://assets.publishing.service.gov.uk/media/64118f0f8fa8f555779ab001/Pro-innovation_Regulation_of_Technologies_Review_-_Digital_Technologies_report.pdf [Accessed February 23, 2025].

[119] Ibid.

[120] HM Government Response to Sir Patrick Vallance's Pro-Innovation Regulation of Technologies Review Digital Technologies. Paras 4-7. (2023) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1142798/HMG_response_to_SPV_Digital_Tech_final.pdf [Accessed February 3, 2024].

[121] The selection of group members lacked transparency (Trapova, 2024). The lack of public documentation of the process is also an issue, with the only available documentation to include the Terms of Reference and the

perspectives on ethics, bias, and cultural impact, which are critical for shaping balanced AI policy. Government should be more inclusive in forming such groups instead of limiting the discussion on such an important code of practice to "AI firms and rights holders".[122]

The working group did not reach consensus on an effective voluntary code.[123] The Culture, Media, and Sport Committee (House of Commons, 2024) expressed concern over the lack of agreement between the creative industries and AI developers regarding creators' consent and compensation concerning the utilisation of their works for AI training purposes. The Committee urged the Government to establish mechanisms that enable creators to enforce their consent and receive equitable compensation when their works are employed by AI systems. Government had previously stated that if the code of practice was not adopted, new legislation could be considered.[124]

More recently, the UK government issued new consultation which continues to focus primarily on the creative industries and AI companies while largely omitting considerations related to cultural heritage.[125] Notably, the consultation proposes a new exception for commercial TDM including an opt-out provision for rights holders, allowing them to exclude their works from AI training datasets. While this aims to protect copyright interests, the possibility of opt-outs raises concerns about potential biases, omissions, and incomplete datasets that could skew and compromise AI research. This is particularly problematic if this provision was to apply in research and heritage contexts, which could become the case in light of the uncertainties regarding the current non-commercial research TDM exception and risk-averse attitudes of heritage stakeholders as discussed in section 2. It is unclear how the proposed new exception will impact the existing TDM exception for non-commercial research, which we believe is not

---

member list: UK IPO. Guidance: The government's code of practice on copyright and AI. (2023) https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai [Accessed February 3, 2024].

[122] HM Government Response to Sir Patrick Vallance's Pro-Innovation Regulation of Technologies Review Digital Technologies. Para. 5. (2023) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1142798/HMG_response_to_SPV_Digital_Tech_final.pdf [Accessed February 3, 2024].

[123] Secretary of State for Science, Innovation and Technology. Consultation Outcome, A pro-innovation approach to AI regulation (CP 1019) (2024) https://assets.publishing.service.gov.uk/media/65c1e399c43191000d1a45f4/a-pro-innovation-approach-to-ai-regulation-amended-governement-response-web-ready.pdf [Accessed June 15, 2024].

[124] Ibid. Para. 6.

[125] UK IPO, Department for Science, Innovation and Technology and Department for Culture, Media and Sport. Open consultation: Copyright and Artificial Intelligence (2024) https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence [Accessed January 31, 2025].

fit for purpose for heritage research, and should be clarified and expanded (as proposed above) before the introduction of any new exception. The boundaries between any new commercial TDM exception and the non-commercial research exception must be carefully and clearly delineated, to resolve existing issues and protect and promote AI research in heritage contexts.

Additionally, Recommendation 13 of the AI Opportunities Action Plan[126] proposes establishing a copyright-cleared British media asset training dataset through partnerships with heritage institutions, which - while framed as a means to advance AI development - effectively commercialises heritage data, reinforcing an industry-driven focus that sidelines broader cultural heritage considerations. The above appear counter to the mission of publicly funded cultural institutions, and as such must be carefully considered in close consultation with heritage stakeholders.[127] Whatever the next policy step is, we advocate that meaningful engagement with the heritage sector in its full breadth and diversity is required.

The reluctance to implement a broader TDM exception, despite early considerations, reflects the dominance of creative industry concerns. Kretschmer *et al*. (2024, p. 112) state that "policy making may be anecdotally driven by examples that surface through lobbying processes or the latest technological applications", the current policy context being dominated by discussions on user-facing generative AI applications such as Chat GPT. But the "real world" (as Kretchmer *et al*. put it) of machine learning is not limited to these more dominant scenarios that attract much attention. We believe this is a crucial point for the need to advocate for further policy work engaging less represented stakeholders, such as those in the wide and diverse heritage sector.

AI and copyright policy discussions tend to focus more on creative industries and AI businesses. This can result in the exclusion, in policy discourses and debates, of the heritage sector. It would be important to understand why such public organisations are not more robustly involved, and how can the language of consultations, call for views and policy reports be

---

[126] Department for Science, Innovation and Technology. AI Opportunities Action Plan, Matt Clifford CBE (2025) https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan#lay-the-foundations [Accessed January 31, 2025].

[127] Government partly agreed with Recommendation 13, stating it "will engage with partner organisations and industry to consider the potential role of government in taking forward this recommendation." Department for Science, Innovation and Technology. AI Opportunities Action Plan Government Response (2025) p. 10 <https://www.gov.uk/government/publications/ai-opportunities-action-plan-government-response/ai-opportunities-action-plan-government-response> [Accessed February 18, 2025].

improved to include the need, more explicitly, for evidence from this sector. It is remarkable that the European Parliament has specifically addressed the intersection of AI and cultural heritage through a dedicated briefing 'Artificial intelligence in the context of cultural heritage and museums', exploring the legal challenges faced by the sector. In contrast, the focus of current efforts in the UK Parliament has been more concentrated on the impact of AI on the creative industries, with less attention given to the cultural heritage sector. Events such as 'Changes and Challenges in Heritage and Open Knowledge' supported by the National Lottery Heritage Fund[128] and 'Roundtable on ICH inventorying, Intellectual Property and Artificial Intelligence'[129] promote important discussions and we believe government should proactively seek such forms of engagement to inform its policy.

CONCLUSION

This paper outlined the current important applications of AI in the UK heritage sector and highlighted the legal challenges, particularly in copyright law, for heritage stakeholders using AI in research and heritage management. High clearance costs and risk-averse attitudes in digitisation projects, alongside the limitations of the TDM exception for non-commercial research, were discussed.

We argued that AI research (and related heritage management) in cultural heritage contexts should benefit from the TDM exception. Licensing may not be feasible or affordable, and can introduce bias and other issues related to dataset completeness and appropriateness. Practical and legal issues, such as ambiguities in the "lawful access" requirement, need to be clarified. Addressing these challenges would support the public interest, alleviate resource constraints, and improve the quality of heritage research and management.

The current UK TDM exception should be interpreted to cover digitisation of physical materials when needed for AI research. The quotation exception can supplement the TDM exception, particularly for sharing excerpts publicly. In view of the problems with the "non-commercial" terminology, expanding the exception to "non-profit" research, and clearly contemplating public-private partnerships should be considered. We also argued that the TDM

---

[128] Naomi Korn et al, <https://www.culturehive.co.uk/digital-heritage-hub/resource/leadership/heritage-open-knowledge/> [Accessed February 19, 2025].
[129] Harriet Deacon, <https://www.hull.ac.uk/work-with-us/research/site-elements/docs/Report-3-June-roundtable-2108.pdf> [Accessed February 19, 2025].

exception should be amended to clearly allow the transfer of copies between research project partners. These issues should be resolved and clarified to support AI heritage research and management, particularly in light of current discussions on including a commercial TDM exception with opt-out.

We highlighted the underrepresentation of cultural heritage stakeholders in UK copyright and AI debates, and the missed opportunity to enrich current policy discussions with important heritage perspectives. This gap underscores the need for more inclusive policy language that invites broader participation beyond AI and creative industries. More empirical research is necessary to identify the copyright issues experienced and the barriers preventing heritage sector involvement in these discussions. We recommend increased proactive government engagement with the heritage sector (including researchers) in AI and copyright policy, building on efforts by bodies such as the National Lottery Heritage Fund. Projects such as *Living With Machines* demonstrate the need to address copyright challenges in heritage research, providing a foundation for future legal and policy reforms.

References

1. Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994).

2. Ahnert R, Griffin E, Ridge M, and Tolfo G. Collaborative Historical Research in the age of big data. Cambridge University Press (2023). Available at: https://www.cambridge.org/core/elements/collaborative-historical-research-in-the-age-of-big-data/839C422CCAA6C1699DE8D353B3A1960D [Accessed February 3, 2024].

3. Aplin T. and Davis J. Intellectual Property Law: Text, Cases and Materials. 4th edition. Oxford: Oxford University Press (2021).

4. Arzomand K, Rustell M, and Kalganova T. From ruins to reconstruction: Harnessing text-to-image AI for restoring historical architectures. Challenge Journal of Structural Mechanics (2024) 10(2), p. 69-85.

5. Authors Guild v Google, Inc, No.13–4829 (2d Cir. 2015).

6. Aslan S, Vascon S, and Pelillo M. Two sides of the same coin: improved ancient coin classification using graph transduction games. Pattern Recognit. Lett (2020) 131, p. 158.

7. Beelen K, Lawrence J, Wilson D C S, Beavan D. "Bias and representativeness in digitized newspaper collections: Introducing the environmental scan" (2023) 38(1) Digital Scholarship in the Humanities 1–22, https://doi.org/10.1093/llc/fqac037

8. Bently L, Sherman B, Gangjee D, and Johnson P. Intellectual Property Law (6th edition). Oxford: Oxford University Press (2022).

9. Bonadio E, Dinev P, and McDonagh L. Can Artificial Intelligence Infringe Copyright? Some Reflections. In: Ryan Abbott, ed. Research Handbook on Intellectual Property and Artificial Intelligence. Cheltenham: Edward Elgar; 2022. p. 245-258.

10. Bordoni L, Mele F, and Sorgente A. (eds), Artificial Intelligence for Cultural Heritage Cambridge Scholars (2016) vii.

11. Bottis M, Papadopoulos M, Zampakolas C, and Ganatsiou P. Text and Data Mining in Directive 2019/790/EU Enhancing Web-Harvesting and Web-Archiving in Libraries and Archives. Open Journal of Philosophy (2019) 9, p. 369.

12. Brown A, Kheria S, Cornwell J, and Iljadica M. Contemporary Intellectual Property: Law and Policy (6th edition). Oxford: Oxford University Press (2023).

13. Butler T. Can a Computer Be an Author?. Hastings Comm. & Ent L J (1982) 4, p. 707.

14. Cantrill-Fenwick R. Spotlight on: AI', *UK Heritage Pulse*, (August 2023). https://heritagepulse.insights-alliance.com/updates/spotlight-on-ai/ [Accessed: February 3, 2024].

15. Cetinic E, Lipic T, and Grgic S. Learning the principles of art history with convolutional neural networks. Pattern Recognit. Lett (2020) 129, p. 56.

16. Chetouani A, Treuillet S, Exbrayat M, and Jesset S. Classification of engraved pottery sherds mixing deep-learning features by compact bilinear pooling. Pattern Recognit. Lett (2020) 131, p. 1.

17. Coleman C N. AI for Digital Research Services: Stanford Libraries AI Initiative Report. Stanford Digital Repository (2019). https://purl.stanford.edu/zy067zc8244 [Accessed April 30, 2024).

18. Collections Trust. Spectrum. Acquisition and accessioning. https://collectionstrust.org.uk/spectrum/procedures/acquisition-and-accessioning-spectrum-5-0/ [Accessed February 23, 2025].

19. Collections Trust. Spectrum. Acquisition and accessioning – suggested procedure. https://collectionstrust.org.uk/resource/acquisition-and-accessioning-suggested-procedure/ [Accessed February 23, 2025].

20. Controller HMSO and Ordnance Survey v Green Amps [2007] EWHC 2755 (Ch).

21. Cranfield University Library Services. Text and Data Mining. https://library.cranfield.ac.uk/text-and-data-mining/copyright [Accessed October 8, 2024].

22. Culture Action Europe. EU Artificial Intelligence Act (March 2024). https://cultureactioneurope.org/news/european-parliament-adopts-the-ai-act/ [Accessed June 15, 2024).

23. Deacon H. Roundtable on ICH inventorying, Intellectual Property and Artificial Intelligence, 3 June 2024 https://www.hull.ac.uk/work-with-us/research/site-elements/docs/Report-3-June-roundtable-2108.pdf [Accessed February 19, 2025].

24. Del Chiaro R, Bagdanov A D, and Del Bimbo A. Webly-supervised zero-shot learning for artwork instance recognition. Pattern Recognit. Lett (2019) 128, p. 420.

25. Department for Science, Innovation and Technology. AI Opportunities Action Plan, Matt Clifford CBE (2025) https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan#lay-the-foundations [Accessed: January 31, 2025].

26. Department for Science, Innovation and Technology. AI Opportunities Action Plan Government Response (2025) https://www.gov.uk/government/publications/ai-opportunities-action-plan-government-response/ai-opportunities-action-plan-government-response [Accessed February 18, 2025].

27. Department for Science, Innovation and Technology. A pro-innovation approach to AI regulation Government response to consultation (CP 1019) (2024). https://assets.publishing.service.gov.uk/media/65c1e399c43191000d1a45f4/a-pro-innovation-approach-to-ai-regulation-amended-governement-response-web-ready.pdf [Accessed June 15, 2024].

28. Department for Science, Innovation & Technology. Policy paper: A pro-innovation approach to AI regulation (2023). https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper [Accessed February 23, 2025].

29. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (2001) Official Journal L167, p. 1.

30. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance.) OJ L 130, 17.5.2019, p. 92–125

31. Drexl J*, Hilty R M, Beneke F, Desaunettes L, Finck M, Globocnik J, Gonzalez Otero B, Hoffmann J, Hollander L, Kim D, Richter H, Scheuerer S, Slowinski P R, Thonemann J. Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective. *Max Planck Institute for Innovation & Competition Research Paper No. 19-13* (2019). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3465577 [Accessed February 3, 2024].

32. Dusollier S. The 2019 Directive on Copyright in the Digital Single Market: Some progress, a few bad choices, and an overall failed ambition (2020) 57(4) CMLR 979.

33. England And Wales Cricket Board Ltd & Anor v Tixdaq Ltd & Anor [2016] EWHC 575 (Ch).

34. ERRIN (2024) "Cultural Heritage & Tourism/ICT WG meeting: Unlocking the Power of AI for Cultural Heritage" https://errin.eu/events/cultural-heritage-tourismict-wg-meeting-unlocking-power-ai-cultural-heritage#:~:text=Through%20advanced%20AI%20algorithms%2C%20data,ensuring%20historical%20integrity%20is%20preserved [Accessed February 18, 2025]

35. European Copyright Society (2017) General Opinion on the EU Copyright Reform Package. https://www.ivir.nl/publicaties/download/ECS_opinion_on_EU_copyright_reform.pdf [Accessed: June 14, 2024].

36. EUIPO, https://www.euipo.europa.eu/en/law/recent-case-law/germany-hamburg-district-court-310-o-22723-laion-v-robert-kneschke [Accessed February 19. 2025].

37. European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf [Accessed May 2, 2024].

38. Europeana. Copyright and Digital Cultural Heritage. https://pro.europeana.eu/page/copyright-and-digital-cultural-heritage [Accessed October 7, 2024].

39. Football Association Premier League Ltd and Others v QC Leisure and Others and Karen Murphy v Media Protection Services Ltd, joined Cases C-403/08 and C-429/08, EU:C:2011:631

40. Geiger C and Izyumenko E. The Constitutionalisation of Intellectual Property Law in the EU and the Funke Medien, Pelham and Spiegel Online Decisions of the CJEU: Progress, But Still Some Way to Go!. IIC (2020) 51(3), p. 282.

41. Geiger C, Frosio G. and Bulayenko O. Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU. Centre for International Intellectual Property Studies Research Paper No. 2019-08 (2019).

42. Geiger C, Hilty R M, Griffiths J and Suthersanen U. 'Declaration: A Balanced Interpretation of the "Three-Step Test" in Copyright Law (2010) 1 Journal of Intellectual Property, Information Technology and E-Commerce Law 119.

43. Getty Images (US) Inc & Ors v Stability AI Ltd [2023] EWHC 3090 (Ch).

44. Gervais D J, Shemtov N, Marmanis H and Zaller Rowland C. The Heart of the Matter: Copyright, AI Training, and LLMs (September 21, 2024). Available at SSRN: https://ssrn.com/abstract=4963711

45. Girbacia F. An analysis of research trends for using AI in cultural heritage. Electronics (2024), 13:18. https://doi.org/10.3390/electronics13183738.

46. Guadamuz A. A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs. GRUR International. 2024 January 13; 73(2):111-127.

47. Hargreaves I. Digital Opportunity: A Review of Intellectual Property and Growth (the Hargreaves Review) (2011). https://researchbriefings.files.parliament.uk/documents/SN06430/SN06430.pdf [Accessed October 5, 2024].

48. Hawkins A and Sichani A M. (2024). 'Data Matters': Report on the Towards a National Collection Discovery Projects focus group on data management, documentation, and archiving practices in digital cultural heritage projects. Towards a National Collection https://doi.org/10.5281/zenodo.14006955

49. Hembt S, Lutzhöft N and Bond T. Long-awaited German judgment by the District Court of Hamburg (Kneschke v. LAION) on the text and data mining exception(s) https://www.twobirds.com/en/insights/2024/germany/long-awaited-german-judgment-by-the-district-court-of-hamburg-kneschke-v-laion [Accessed February 19. 2025].

50. Heritage Digital. Heritage organisations & exceptions to copyright. 2021. https://naomikorn.com/wp-content/uploads/2021/07/Guide-to-Exceptions_Final-.pdf [Accessed: October 7, 2024].

51. HM Government. HM Government Response to Sir Patrick Vallance's Pro-Innovation Regulation of Technologies Review Digital Technologies. (2023) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1142798/HMG_response_to_SPV_Digital_Tech_final.pdf [Accessed February 3, 2024].

52. HM Government. Pro-innovation Regulation of Technologies Review: Digital Technologies. Report presented by Sir Patrick Vallance. (2023) https://assets.publishing.service.gov.uk/media/64118f0f8fa8f555779ab001/Pro-innovation_Regulation_of_Technologies_Review_-_Digital_Technologies_report.pdf [Accessed February 23, 2025].

53. HM Government. Modernising Copyright: A Modern, Robust and Flexible Framework. Government Response to Consultation on Copyright Exceptions and Clarifying Copyright Law (2012) 37. https://data.parliament.uk/DepositedPapers/Files/DEP2012-1936/ModernisingCopyrightGovernmentresponse.pdf [Accessed October 6, 2024].

54. House of Commons, Culture, Media and Sport Committee. Creator Remuneration. (2024) https://committees.parliament.uk/publications/44143/documents/219382/default/ [Accessed May 1, 2024].

55. House of Lords Communications and Digital Committee. Chapter 2: A Digital Future. (2023) https://publications.parliament.uk/pa/ld5803/ldselect/ldcomm/125/12505.htm#_idTextAnchor 019 [Accessed February 3, 2024].

56. Hudson E. Drafting Copyright Exceptions: From the Law in Books to the Law in Action. Cambridge University Press (2020), p. 154.

57. Hugenholtz B, and Quintais J P. Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?. IIC (2021) 52(9), p. 1190.

58. Iglesias Portela M, Shamuilia S, and Anderberg A. Intellectual Property and Artificial Intelligence – A literature review, Publications Office of the European Union, EUR30017EN (2019) https://publications.jrc.ec.europa.eu/repository/handle/JRC119102 [Accessed February 3, 2024].

59. Infopaq International A/S v Danske Dagblades Forening (2012) Case no. C-5/08; ECLI:EU:C:2009:465. *European Court Reports,* I, 06569.

60. Jacques S. On the Wax or on the Wane? The Influence of Fundamental Rights in Shaping Exceptions and Limitations. In Rosati E. (ed.) The Routledge Handbook of EU Copyright Law. Routledge (2021).

61. Knowledge Integration. Introducing the Museum Data Service. https://www.k-int.com/introducing-the-museum-data-service/ [Accessed October 7, 2024].

62. Korn N. In from the Cold: An assessment of the scope of 'Orphan Works' and its impact on the delivery of services to the public. JISC (2009) https://naomikorn.com/wp-content/uploads/2020/09/SCA_CollTrust_Orphan_Works_v1-final.pdf.

63. Korn N, Waterman S, and Naomi Korn Associates. Changes and Challenges in Heritage and Open Knowledge https://www.culturehive.co.uk/digital-heritage-hub/resource/leadership/heritage-open-knowledge/ [Accessed February 19, 2025].

64. Kretschmer M, Margoni T, and Oruc P. Copyright Law and the Lifecycle of Machine Learning Models. IIC (2024) 55, p. 110-138.

65. Lampert C and Lapworth E. What do we mean by "Collections As Data" (CAD)? https://www.library.unlv.edu/whats-new-special-collections/2020/2020-03/what-do-we-mean-collections-data-cad-cory-lampert-emily

66. Levendowski A. How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem Washington Law Review (2018) 93, 579.

67. Mammen C E, and Richey C. AI and IP: Are Creativity and Inventorship Inherently Human Activities?. FIU L Rev (2020) 14, p. 275.

68. Margoni T and Kretschmer M. A deeper look into the EU Text and Data Mining exceptions: Harmonisation, data ownership, and the future of technology. GRUR International (2022) 71(8), p. 685.

69. Matas A. AI 'opt-outs': should cultural heritage institutions (dis)allow the mining of cultural heritage data? https://pro.europeana.eu/post/ai-opt-outs-should-cultural-heritage-institutions-dis-allow-the-mining-of-cultural-heritage-data [Accessed February 19, 2025].

70. McDonald D and Kelly U. The Value and Benefit of Text Mining. JISC (2012) https://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining [Accessed February 23, 2025].

71. Metzig C, Gould M, Noronha R, Abbey R, Sandler M, and Colijn C. Classification of origin with feature selection and network construction for folk tunes. Pattern Recognit. Lett (2020) 133, p. 356.

72. Miller A R. Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?. Harv L Rev (1993) 106, p. 977.

73. Mitchell K. V&A Blog. The Algorithmic Gesture: Sougwen Chung's MEMORY (2022). https://www.vam.ac.uk/blog/digital/the-algorithmic-gesture-sougwen-chungs-memory?srsltid=AfmBOoqdmjXhwgiHYop09besEAR-7i5nWI7NEsXnTsAqhoC0VM4YOM-i&doing_wp_cron=1728552593.7939109802246093750000 [Accessed February 1, 2025]

74. Murphy O and Villaespesa E. AI: A Museum Planning Toolkit, The Museums + AI Network, Goldsmiths University of London (2020). https://themuseumsainetwork.files.wordpress.com/2020/02/20190317_museums-and-ai-toolkit_rl_web.pdf [Accessed February 3, 2024].

75. Murray-Rust P. The Right to Read is the Right to Mine. Open Knowledge; (2012). https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/ [Accessed October 7, 2024].

76. Network of European Museum Organisations. Museums and Artificial Intelligence (March 2024). https://www.ne-mo.org/advocacy/our-advocacy-themes/museums-and-ai-conference/summary [Accessed June 15, 2024].

77. Nguyen K C, Nguyen C T, and Nakagawa M. Nom document digitalization by deep convolution neural networks. Pattern Recognit. Lett (2020) 133, p. 8.

78. Ostertag C and Beurton-Aimar M. Matching ostraca fragments using a Siamese neural network. Pattern Recognit. Lett (2020) 131, p. 336.

79. Paiva P V, Cogima C, Dezen-Kempter E, and Carvalho M. Historical building point cloud segmentation combining hierarchical watershed transform and curvature analysis. Pattern Recognit. Lett (2020) 135, p. 114.

80. Pasikowska-Schnass M and Lim Y-S. European Parliament Briefing: Artificial intelligence in the context of cultural heritage and museums (PE 747.120 – May 2023). https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747120/EPRS_BRI(2023)747120_EN.pdf [Accessed April 30, 2024].

81. Patton S. AI Meets Archives: The Future of Machine Learning in Cultural Heritage (CLIR, October 21, 2024) https://www.clir.org/2024/10/ai-meets-archives-the-future-of-machine-learning-in-cultural-heritage/ [Accessed 18 February 2025]

82. Pavis M. Digital Heritage Leadership Briefing: Artificial Intelligence supported by The National Lottery Heritage Fund. (2023) https://www.heritagefund.org.uk/sites/default/files/media/attachments/Digital%20Heritage%20Leadership%20Briefing%20-%20Artificial%20Intelligence_English.pdf [Accessed April 30, 2024].

83. Pelham GmbH and Others v Ralf Hütter and Florian Schneider-Esleben (2019) Case no. C-476/17; ECLI:EU:C:2019:624.

84. Ramalho A. Will Robots Rule the (Artistic) World?: A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems. Journal of Internet Law (2017) 21(1), 1.

85. Rendas T. Spiegel Online: Do Copyright Exceptions and Fundamental Rights Make Easy Bedfellows?. JIPLP (2018) 13(1), 10.

86. Rahmatian A. Originality in UK Copyright Law: The Old "Skill and Labour" Doctrine Under Pressure. IIC 44, 4–34 (2013). https://doi.org/10.1007/s40319-012-0003-4

87. Ridge M. Keynote video 'Evolutionary Innovations: Collections as Data in the AI era' for Making Meaning 2024 <https://www.miaridge.com/keynote-evolutionary-innovations-collections-as-data-in-the-ai-era-for-making-meaning-2024/> [Accessed February 19, 2025].

88. Robert Kneschke v LAION gemeinnütziger e.V. (Case No. 310 O 227/23), Regional Court (Landgericht) of Hamburg.

89. Romanengo C, Biasotti S, and Falcidieno B. Recognising decorations in archaeological finds through the analysis of characteristic curves on 3D models. Pattern Recognit. Lett (2020) 131, p. 405.

90. Rosati E. Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity. Asia Pacific Law Review (2019) 27(2), p. 198.

91. Secretary of State for Science, Innovation and Technology. A pro-innovation approach to AI regulation (Cm 815) Presented to Parliament by the Secretary of State for Science, Innovation and Technology by Command of His Majesty on 29 March 2023. (2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#executive-summary [Accessed February 3, 2024].

92. Secretary of State for Science, Innovation and Technology (2024) Consultation Outcome, A pro-innovation approach to AI regulation (CP 1019) Presented to Parliament by the Secretary of State for Science, Innovation and Technology by Command of His Majesty on 6 February 2024. https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response [Accessed May 1, 2024].

93. Sheikh H, Prins C, and Schrijvers E. Artificial Intelligence: Definition and Background. In: Mission AI. Research for Policy. Springer (2023) p. 15-41.

94. Skoki A, Ljubic S, Lerga J, and Stajduhar I. Automatic music transcription for traditional woodwind instruments sopele., Pattern Recognit. Lett (2019) 128, p. 340.

95. Stratton B. Seeking New Landscapes. A Rights Clearance Study in the Context of Mass Digitisation of 140 Books Published between 1870 and 2010. The British Library (2011). https://www.ifla.org/wp-content/uploads/files/assets/hq/topics/exceptions-limitations/documents/bl_rights_clearance_study.pdf [Accessed October 5, 2024].

96. Strowel A and Ducato R. Artificial Intelligence and Text and Data Mining: A Copyright Carol. In Rosati E, The Routledge Handbook of EU Copyright Law. Routledge (2021).

97. Synodinou T E. Lawfulness for Users in European Copyright Law: Acquis and Perspectives. (2019) 10 JIPITEC 20.

98. The Berne Convention for the Protection of Literary and Artistic Works (adopted in 1886, entered into force 5 December 1887) 1161 UNTS 30 (Berne Convention).

99. THJ Systems Limited & Anor v Daniel Sheridan & Anor [2023] EWCA Civ 1354.

100. Tolfo G. Vane O, Beelen K, Hosseini K, Lawrence J, Beavan D, and McDonough K. Hunting for Treasure: Living with Machines and the British Library Newspaper Collection. In: Estelle Bunout, Maud Ehrmann and Frédéric Clavert, editors. Digitised Newspapers – A New Eldorado for Historians?: Reflections on Tools, Methods and Epistemology, Berlin, Boston: De Gruyter Oldenbourg (2023), pp. 23-46. https://doi.org/10.1515/9783110729214-002

101. Torremans P. Presenting Best Practices for Copyright and Trademarks in Common Law Countries for Cultural Heritage and Museum Collections. *WIPO Workshop on IP for Libraries/Museums.* Bibliotheca Alexandrina, Alexandria, Egypt, 11 September 2019.

102. Torremans P. The protection of cultural heritage by copyright and related rights. In: Stamatoudi I, editor. Research Handbook on Intellectual Property and Cultural Heritage. Cheltenham, UK: Edward Elgar Publishing (2022), pp. 38-55.

103. Trapova A. UK's Short-Lived Dream for a Code of Practice on GenAI and Copyright Law. Kluwer Copyright Blog (2024). https://copyrightblog.kluweriplaw.com/2024/02/28/uks-short-lived-dream-for-a-code-of-practice-on-genai-and-copyright-law/ [Accessed June 15, 2024].

104. Tsigkas G, Sfikas G, Pasialis A, Vlachopoulos A, and Nikou C. Markerless detection of ancient rock carvings in the wild: rock art in Vathy, Astypalaia., Pattern Recognit. Lett (2020) 135*, p. 337.

105. UCL Library Services. The text and data mining (TDM) exception. https://www.ucl.ac.uk/library/learning-teaching-support/ucl-copyright-advice/copyright-depth/text-and-data-mining-tdm-exception [Accessed: October 8, 2024].

106. UK Copyright, Designs and Patents Act 1988.

107. UK Intellectual Property Office, Department for Science, Innovation and Technology and Department for Culture, Media and Sport. Open consultation: Copyright and Artificial Intelligence (2024) https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence [Accessed January 31, 2025].

108. UK Intellectual Property Office. Exception for copying of works for use by text and data analytics. Impact Assessment. (2012) https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/consult-ia-bis0312.pdf [Accessed October 5, 2024].

109. UK Intellectual Property Office. Exceptions to copyright: Research. (2014) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf [Accessed June 14, 2024].

110. UK Intellectual Property Office (2021) Consultation stage impact assessment on Artificial Intelligence and Intellectual Property. Impact Assessment. https://assets.publishing.service.gov.uk/media/617bfda2d3bf7f55fc098f73/impact-assessment.pdf [Accessed October 5, 2024].

111.   UK Intellectual Property Office (2014) The Technical Review of Draft Legislation on Copyright Exceptions: Government Response (HM Government, Intellectual Property Office, March 2014, DPS/IP B900-03/14). https://webarchive.nationalarchives.gov.uk/ukgwa/20140603083549/http://www.ipo.gov.uk/response-copyright-techreview.pdf [Accessed October 8, 2024].

112.   UK Intellectual Property Office. Artificial Intelligence and IP: Copyright and Patents Consultation Outcome. (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents [Accessed February 3, 2024].

113.   UK Intellectual Property Office. Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation (2022) https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation [Accessed October 8, 2024].

114.   UK Intellectual Property Office. Guidance: The government's code of practice on copyright and AI. (2023) https://www.gov.uk/guidance/the-governments-code-of-practice-on-copyright-and-ai [Accessed February 3, 2024].

115.   UK Intellectual Property Office. Government response to call for views on artificial intelligence and intellectual property. Consultation outcome (2021) https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property [Accessed May 1, 2024].

116.   UK Parliament. Copyright: EU Action. (2020) https://questions-statements.parliament.uk/written-questions/detail/2020-01-16/4371 [Accessed October 6, 2024].

117.   Vuckovic R M, Kanceljak I, and Juric M. Cultural Heritage Institutions during and after the Pandemic - The Copyright Perspective. ECLIC (2021) 5, p. 379.

118.   Wallace A. 21 for 2021: Digital heritage and the public domain. CREATe 2022. https://www.create.ac.uk/blog/2022/01/07/21-for-2021-digital-heritage-and-the-public-domain/ [Accessed: October 5, 2024].

119.   Wallace A. Surrogate Intellectual Property Rights in the Cultural Sector. Journal of Law, Technology and Policy (2023), 303.

120.   Walter v Steinkopff [1892] 3 Ch 489.

121.   Westenberger P, Digital Culture, Copyright and the Orphan Works Issue: a view from Brazil. In Fortes, Boratti et al, editors, Law and Policy in Latin America: Transforming Courts, Institutions, and Rights. Palgrave Macmillan. (2017), p. 293-309.

122.   WIPO Copyright Treaty (WCT) (adopted in December 20, 1996).

123.   WTO, Panel Rep. of 15 June 2000, WT/DS160/R.

124.   Wu A J. From Video Games to Artificial Intelligence: Assigning Copyright Ownership to Works Generated by Increasingly Sophisticated Computer Programs. AIPLA Q L (1997) 25, p. 131.

125.   Yanisky-Ravid S. Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era-the Human-Like Authors Are Already Here-A New Model. MICH ST L REV (2017), p. 659.