GRUR International, XX(XX), 2025, 1–16

https://doi.org/10.1093/grurint/ikaf093
Opinion

MARTIN KRETSCHMER[I] / BARTOLOMEO MELETTI[II] / LIONEL BENTLY[III] / GABRIELE CIFRODELLI[IV] / MAGALI EBEN[V] / KRISTOFER ERICKSON[VI] / ALINE IRAMINA[VII] / ZIHAO LI[VIII] / LUKE MCDONAGH[IX] / EMMA PEROT[X] / LUIS PORANGABA[XI] / AMY THOMAS[XII]

Copyright and AI in the UK: Opting-In or Opting-Out?

The interface between copyright law and artificial intelligence (AI) is currently the object of global attention. The UK government's recent public consultation on policy to ensure that 'the UK's legal framework for AI and copyright supports the UK creative industries and AI sector together' is an example of this wider interest, as countries engage in a form of regulatory competition for the most attractive environment for AI development. The option endorsed by the UK government in the consultation document remains close to the EU model, proposing 'a data mining exception which allows right holders to reserve their rights, supported by transparency measures'. We argue that this approach is a missed opportunity for a more straightforward innovation policy that avoids the problems of rights reservation. Opt-outs from training are difficult to implement technically, they increase costs and create barriers to market entry. Instead, we suggest that there should be much clearer scope for permitted research before market entry within the traditional opt-in framework (avoiding the EU split between commercial and non-commercial research). This should be combined with transparency obligations (triggering potential licensing) and an equitable remuneration provision that enables creatives (authors, artists, performers) to receive a share of licensing revenues negotiated between AI developers and intermediaries (publishers, producers). We review evidence on current licensing practices and consider related legal issues, including data protection, image rights and the *sui generis* database right.

I. Introduction

In December 2024, the UK government launched a public consultation[1] 'to find the right balance between encouraging innovation in AI in the UK while ensuring protection for creators and our vibrant creative industries' (para. 52). The objective was to 'enable licensing agreements, ensuring right holders are remunerated where appropriate' (para. 72) while achieving a UK copyright environment that is 'at least as attractive as in competing jurisdictions'.[2] This article draws on the submission of the CREATe[3] Centre at the University of Glasgow to the consultation. While focussing on the UK, the issues involved in balancing the interest of different groups are of wide interest in a globally competitive environment where digital technology is not easily contained by jurisdictional borders.

The UK government's favoured proposal to achieve a balance between AI technology innovators, copyright owners, creators and researchers is controversial, involving 'a data mining exception which allows right holders to reserve their rights, supported by transparency measures' (Option 3). This has been reframed by prominent interests as a choice between opt-in and opt-out, between creative industries and big tech. We contest that this dichotomy is misleading and could lead to poor policy.

I Professor of Intellectual Property Law, CREATe Centre, University of Glasgow, United Kingdom.
II Senior Lecturer in Law, CREATe Centre, University of Glasgow, United Kingdom.
III Herchel Smith Professor of Intellectual Property Law, University of Cambridge, United Kingdom.
IV Research and Teaching Assistant, CREATe Centre, University of Glasgow, United Kingdom. ORCID: 0000-0003-4552-4652
V Senior Lecturer in Competition Law, CREATe Centre, University of Glasgow, United Kingdom. ORCID: 0000-0003-1069-7609
VI Professor of Social Data Science, CREATe Centre, University of Glasgow, United Kingdom.
VII Lecturer in Law, CREATe Centre, University of Glasgow, United Kingdom.
VIII Lecturer in Technology Law, CREATe Centre, University of Glasgow, United Kingdom.
IX Associate Professor in Law, LSE, London, United Kingdom. ORCID: 0000-0003-2085-5404
X Lecturer in Law, University of the West Indies, Kingston, Jamaika.
XI Senior Lecturer in Intellectual Property Law, CREATe Centre, University of Glasgow, United Kingdom.
XII Lecturer in Intellectual Property and Information Law, CREATe Centre, University of Glasgow, United Kingdom.
Lead drafters were Martin Kretschmer and Bartolomeo Meletti. Specific contributions included Lionel Bently (Analysis 3.4: Remuneration), Gabriele Cifrodelli and Magali Eben (Analysis 3.5: Terms of use), Kris Erickson (Evidence 2.3: Open source and local computing), Aline Iramina (Analysis 3.3: Transparency), Zihao Li (Analysis 3.3: Transparency/open source and 3.6: Image rights/data protection), Emma Perot, Luis Porangaba and Luke McDonagh (Analysis 3.6: Image rights), Amy Thomas (Evidence 2.2: Licensing deals and earnings of creatives). This work was supported by the Arts and Humanities Research Council AHRC (grant number AH/Y007662/1).

1 UK Intellectual Property Office, *Copyright and Artificial Intelligence* (17 December 2024) <https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence> accessed 28 May 2025.

2 'Matt Clifford report' (endorsed by the government), Recommendation 24: Reform the UK text and data mining regime so that it is at least as competitive as the EU.

3 CREATe is the Centre for Regulation of the Creative Economy, based at the University of Glasgow, United Kingdom. It was established as a national centre for research on copyright in 2012, initially with funding from AHRC, EPSRC and ESRC. Since 2024, CREATe is funded as UK research infrastructure by the AHRC.

The debate about the status of copyright protected materials in the training of AI models needs to take note of complex evidence (which is available). Empirical evidence can help correct frequent misconceptions, including the impact of future training data sets compared to those already possessed by incumbent firms, as well as the realistic prospects for the UK as a medium-size jurisdiction to exert regulatory pressure in a global marketplace for AI training.

The target of much of the ire, the big tech companies of Alphabet (Google), Amazon, Meta (Facebook), Microsoft, and OpenAI, already have much of the training content they need. They have collected data from their existing services and users (likely with 'consent' under their terms of service) and from 'publicly available' internet scrapes (which may include unlawful sources). Since the release of ChatGPT in November 2022 and rising public awareness, big tech also has engaged in numerous deals with content providers. In many parts of the AI ecosystem we have already moved to a licensing economy. It is no longer a choice between opt-in and opt-out. Powerful genAI models exist and will not disappear.

While the most dominant lobbies (on both sides) present the choices as those of regulatory competition, with the potential for making the UK a more attractive place to invest, in reality, the regulatory environment for R&D in AI is global. With respect to copyright law, the EU is developing opt-out and transparency standards; in the US, litigation will result in many settlements and an understanding that some, but not all scraping and training activities are permitted under the copyright defence of 'fair use'; Chinese firms are releasing (partially) open-source models of uncertain data provenance, such as DeepSeek's R1. The ease of cross-border transfer of code and the legal risk in major jurisdictions make it likely (and desirable) that solutions will not be unique to one country.

This does not mean that the UK's choices do not matter, but they will have more modest effects than the rhetoric suggests. We argue that focus should be on understanding the effects of copyright regulation on research and development (R&D) activities and market entry, and crucially on understanding the effects of copyright law on the distribution of money that is already flowing. We argue that interventions should focus on:

- Removing obstacles for scientific research and education, recognising the specific imperatives of the humanities, social sciences and cultural sectors;
- Improving and clarifying the licensing environment, so that access to quality content for R&D becomes easier; this must include transparency requirements;
- Improving the standing and revenues of creatives (authors, artists, performers).

Presenting the current UK proposals as 'the most damaging legislation for the creative industries in decades'[4] is misleading. The interests of creatives (authors, artists, performers) in particular should not be bundled with the owners of large back-catalogues of rights. Legislators also need to keep in mind that copyright law has horizontal effects on non-entertainment sectors, including the science, technical, engineering and mathematical (STEM) disciplines, as well as social sciences and the humanities.

The structure of the article is as follows. Part II reviews the existing evidence on the relationship between copyright law and AI training activities and the potential effects of regulatory reform, particularly in relation to licensing opportunities and the earnings situation of creatives. Part III presents the legal analysis of six key issues: opt-outs, the exception for data mining, transparency requirements, remuneration of creatives, terms of use, and image rights. Finally, Part IV concludes.

II. Existing evidence

From the perspective of innovation policy, the UK seeks to maintain its attractiveness as a destination for AI development and the quality of its research environment, which includes internationally recognised universities, a well-supported research sector and a culture of openness.

The current text and data mining (TDM) exception for non-commercial research (the UK TDM exception) Sec. 29A of the UK Copyright, Designs and Patents Act 1988 (CDPA) is not working as intended. The exception was introduced in 2014 in response to the Hargreaves Review.[5] Hargreaves warned of the need for a future-proof copyright regime and the horizontal effects of copyright beyond its underlying aims, including on scientific and medical research.[6] The government at the time shared Hargreaves' concerns[7] and enacted limited copyright reform skewed by compromise between conflicting industry and civil society claims. The TDM exception was originally implemented with restrictions imposed by EU law that now can be overcome (see section III below). While the existing exception can be interpreted as permitting certain AI training activities, it does not cover the whole lifecycle of AI models and systems.[8] Together with other legal regimes, it

4 Ed Newton-Rex, 'The UK's AI & copyright proposals would irreparably harm the country's creators' (17 December 2024) <https://ednewtonrex.medium.com/the-uks-ai-copyright-proposals-would-irreparably-harm-the-country-s-creators-f20060490f3e> accessed 28 May 2025.

5 Ian Hargreaves, 'Digital Opportunity: A Review of Intellectual Property and Growth' (May 2011) <https://assets.publishing.service.gov.uk/media/5a796832ed915d07d35b53cd/ipreview-finalreport.pdf> accessed 28 May 2025.

6 One of the examples provided by Hargreaves was a set of 1,000 journal papers from the first half of the 20th century describing malaria in indigenous peoples, soldiers, and details of malaria therapy, which was unavailable to researchers because of copyright restrictions; ibid 46-47.

7 'Of particular concern both to the [Hargreaves] Review and to the Government are missed opportunities beyond the core IP-owning industries: vital medical research held up, cultural and commercially useful works locked away for generations and crumbling in archives for want of an owner to give permission for their use, and great business ideas that can't be turned into successful, growing businesses.' HM Government, 'The Government Response to the Hargreaves Review of Intellectual Property and Growth' (2011) 2 <https://assets.publishing.service.gov.uk/media/5a7970f4ed915d042206873c/11-1199-government-response-to-hargreaves-review.pdf> accessed 28 May 2025.

8 Martin Kretschmer, Thomas Margoni and Pinar Oruc, 'Copyright Law and the Lifecycle of Machine Learning Models' (2024) 55 IIC 110-38 <https://doi.org/10.1007/s40319-023-01419-3> accessed 28 May 2025. See also Séverine Dusollier and others, 'Copyright and Generative AI: Opinion of the European Copyright Society' (2025) 16 JIPITEC 121 <https://www.jipitec.eu/jipitec/article/view/424> accessed 28 May 2025.

creates uncertainties at all stages of AI research and development.

The government's ambition to 'responsibly unlock both public and private data sets to enable innovation by UK startups and researchers and to attract international talent and capital' is commendable but requires further thinking and an appropriate legal framework.[9] To create and share high-quality datasets, in addition to incentives and rewards (Recommendation 12), researchers need a clearer and more flexible legal regime that enables data collection and sharing of outputs. Establishing 'a copyright-cleared British media asset training data set … through partnering with bodies that hold valuable cultural data like … the British Library and the BBC' (Recommendation 13) is particularly challenging, especially if done through traditional rights clearance and licensing. The British Library and the BBC own copyright in a fraction of their collections and archives. The copyright status of most third-party items in cultural heritage collections is complex and difficult to ascertain.

There is ample evidence that rights clearance and existing licensing schemes are not suitable solutions to enable data collection and unlock cultural data.[10] In its current form, neither is the UK TDM exception. To provide an example, the Visual Geometry Group (VGG, University of Oxford), one of the world-leading academic groups in computer vision, report themselves to benefit from the TDM exception to use third-party datasets for research purposes. The exception is particularly helpful when licensing is not an option because the copyright status of the data is either complex or uncertain, such as in large, publicly available datasets (e.g. ImageNet)[11] or data scraped from the web.[12] However, in its current form, Sec. 29A substantially limits the VGG's ability to share new datasets they create, which would serve as research benchmarks for the global R&D community. In the social sciences, deep uncertainties limit research activities at all stages of data collection, processing, analysis, and sharing of outputs. These change depending on the type and purpose of the research, but the fundamental legal issues affecting availability and usability of the data – including those listed below – are the same.[13] In the arts and humanities, an illustrative example of the inability of existing licensing schemes and the current TDM exception to unlock cultural data is the Living with Machines project (2018-2023).[14] Despite a substantial budget (GBP 9,082,131), a ground-breaking partnership and the presence of rights expertise within the project, access to and use of already digitised data for non-commercial research was very problematic.[15] Notwithstanding the TDM exception, considerable public funding is being spent to negotiate, access, transfer and pre-process vast amounts of data on a project-by-project basis. Only to then delete the data after the project end date based on uncertainty or access agreements, for no-one's benefit.

Evidence on the experience of data-driven research projects points to the following issues of the existing TDM exception in UK law. First, the existing exception does not reflect current research practices. By only covering the 'making' of copies [Sec. 29A(1)] for non-commercial research purposes and prohibiting the unauthorised transfer of such copies [Sec. 29A(2)(a)], the exception does not allow knowledge transfer in partnerships between different academic institutions, let alone in public-private partnerships. Without authorisation, datasets obtained or created under the existing exception cannot be shared lawfully with the wider research community. Second, while in principle the exception cannot be overridden by contract [Sec. 29A(5)], in practice for various reasons contractual arrangements often regulate access to and the computational use of data for research. The 'lawful access' requirement is ambiguous and open to conservative interpretations, leading to risk-averse behaviour and rightsholders' decisions shaping data-driven research. Third, unlike the EU TDM exception for scientific research, there are no data retention safeguards in the UK exception, with implications for the reproducibility and replicability of research over time. Fourth, the relationship between technological protection measures (TPMs) and exceptions, including the TDM exception, remains problematic.[16] Finally, copyright law is not the only relevant legal regime in relation to data scraping and mining activities. The *sui generis* database right, data protection law, contract law, confidentiality, and competition law pose additional risks and uncertainties.[17] Some of these are addressed in the Positions section below.

9 'AI Opportunities Action Plan', Recommendations 7-13.

10 See Victoria Stobo and others, '"I should like you to see them some time": an empirical study of copyright clearance costs in the digitisation of Edwin Morgan's scrapbooks' (2018) 74(3) Journal of Documentation 641-67.

11 ImageNet is a project led by Stanford University and Princeton University which provides researchers around the world with high-quality image data for training large-scale object recognition models. Researchers and educators can request access to the dataset for non-commercial research and educational purposes: <https://image-net.org/download> accessed 28 May 2025.

12 The University of Oxford holds all the traditional Higher Education collective copyright licences <https://www.bodleian.ox.ac.uk/ask/how-to-guides/copyright/copyright-licences> accessed 28 May 2025 – but none of them covers the content available on online platforms such as YouTube or data scraping activities. Clearing rights on millions of items (eg the ImageNet dataset) is not realistic either.

13 See the case study on machine learning for scientific purposes, in the context of research on regional short-term letting markets, in Kretschmer, Margoni and Oruc (n 8).

14 'Living with Machines, AH/S01179X/1' <https://gtr.ukri.org/projects?ref=AH%2FS01179X%2F1> accessed 25 May 2025. Project partners included The Alan Turing Institute, the British Library, and Universities of Cambridge, East Anglia, Exeter, Queen Mary University of London, and King's College London.

15 So much so that the researchers decided to dedicate a whole chapter of their book *Collaborative Historical Research in the Age of Big Data* to the challenges to data access and use posed by UK copyright law, contractual arrangements, digitisation funding and related policies. Ruth Ahnert and others, *Collaborative Historical Research in the Age of Big Data* (CUP 2023) <https://www.cambridge.org/core/elements/collaborative-historical-research-in-the-age-of-big-data/839C422CCAA6C1699DE8D353B3A1960D> accessed 25 May 2025.

16 A survey of European research libraries revealed that many had faced impediments to carrying out TDM, machine learning and other bibliometric scraping work in support of research. See Kristofer Erickson and Victoria Stobo, 'Evidence on Technological Protection Measures: Impacts for Researchers, Libraries and Archives' (2024) <https://zenodo.org/records/14168677> and <https://www.create.ac.uk/project/public-domain/2024/06/16/evidence-on-technological-protection-measures-impact-on-research-education-and-preservation/> both accessed 28 May 2025.

17 See Sheona Burrow, 'The Law of Data Scraping: A review of UK law on text and data mining' (2021) CREATe Working Paper 2021/2 <https://zenodo.org/records/4635759#.YGLfnD8o9hE> accessed 28 May 2025.

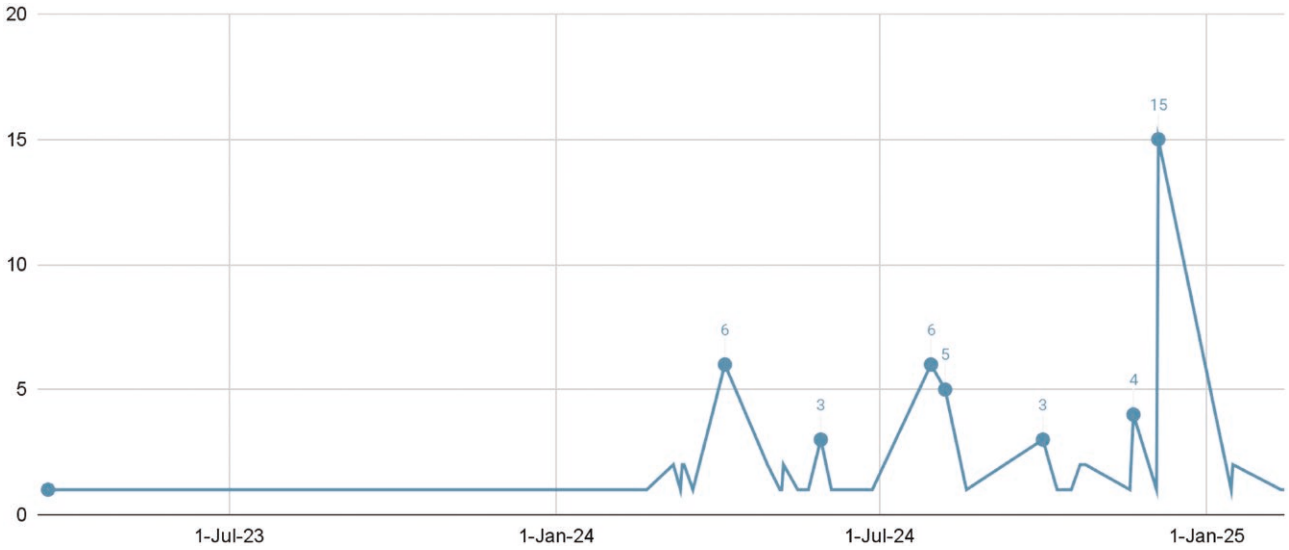

Figure 1 Number (83) of known commercial agreements, by estimated date of conclusion

These issues have provided an opening for collecting societies to offer additional workplace licensing of TDM activities to universities and to the National Health Service (NHS).[18] Increasing costs and bureaucratic hurdles give researchers licensed access to a limited repertoire of content. Licensing can play a role, but the UK needs a more forward-looking approach to copyright law. The position outlined below in the context of the data mining exception would be an important step towards unlocking data assets and reflect the UK's ambition as a destination for research and development.

1. Licensing deals for AI training data and earnings of creatives

Licensing deals for data access for model training are accelerating, with rapidly increasing revenues. These are likely to continue, regardless of the UK's opt-in/opt-out choices, because of the need of AI developers to access quality content and because there are risks of non-compliance with the laws of other major markets. At the same time, earnings of authors, artists and performers are declining.[19] In this section, we first review the evidence on the evolving licensing economy, followed by a summary of income trends for primary creators or creatives.[20]

a) Model training and data access deals 2022-2024

We analysed data based on 83 'known commercial agreements' between content providers (licensors) and AI developers (licensees).[21] These figures do not include agreements which are speculated or anticipated,[22] nor 'non-commercial' agreements between AI developers and their users (e.g., through terms of service),[23] nor openly licensed resources, like Wikipedia. Data were collected based on publicly available information and reporting from parties to the agreement or third-party news reports.[24] The specifics of these agreements are often vague, presumably due to the existence of confidentiality clauses; however, we have confirmed that 94% of the agreements surveyed included an explicit content licence from the licensor to the licensee.[25] Where a term cannot be confirmed, this is listed as 'not specified' – this does not mean the clause does not exist, but rather that it cannot be publicly verified.

Figure 1 shows that the frequency of known commercial agreements has increased since mid-2024. Most of the spikes (indicated with circular data points) below can be attributed to agreements made between a single AI developer and multiple content providers at a time. For example, AI provider Perplexity announced agreements with 11 different content providers on 5 December 2024.

18 See for example 'TDM licensing solution' <https://cla.co.uk/cla-products/business-licence/tdm-licensing/> accessed 28 May 2025.

19 See 'Creators' Earnings Hub' <https://www.create.ac.uk/internal/2024/06/16/creators-hub/> accessed 28 May 2025.

20 Definitional note: While copyright law distinguishes between rights first owned by authors and neighbouring rights (such as for performers), the cultural economics literature often uses the term primary creators to refer to all creatives (who may include artists, musicians, actors or designers whose legal interests may straddle multiple rights). We try to differentiate where we can between legal interests and economic functions.

21 The data has been published on the CREATe Centre website where also the full spreadsheet is linked: <https://www.create.ac.uk/blog/2025/02/24/the-ai-licensing-economy/> accessed 28 May 2025.

22 For example, the speculated deal between Automattic (owner of Tumblr and WordPress) and Midjourney and OpenAI <https://wptavern.com/automattic-faces-scrutiny-over-ai-access-policy> accessed 28 May 2025.

23 For example, users of X agree to terms of service which permit the use of their data for training Grok, an AI tool also owned and developed by X <https://x.com/en/tos> accessed 28 May 2025.

24 In compiling this resource, we consulted trackers from Scholarly Kitchen (scholarly content) <https://scholarlykitchen.sspnet.org/2024/10/15/licensing-scholarly-content-llms/>, Press Gazette (news content) <https://pressgazette.co.uk/platforms/news-publisher-ai-deals-lawsuits-openai-google/>, and CB Insights (a 'comprehensive' list, including audiovisual content providers) <https://www.cbinsights.com/research/ai-content-licensing-deals/> all accessed 28 May 2025. Keyword searches were also conducted to check for completeness with comprehensive Boolean search strings. example: ('licensing agreement' or 'content licensing deal' or 'partnership' or 'contract' or 'rights agreement' or 'data licensing' or 'licensing deal') and ('AI company' or 'artificial intelligence firm').

25 The remainder (Fellowships 4% and Partnerships 2%) are broadly expressed and include implied licences of content.

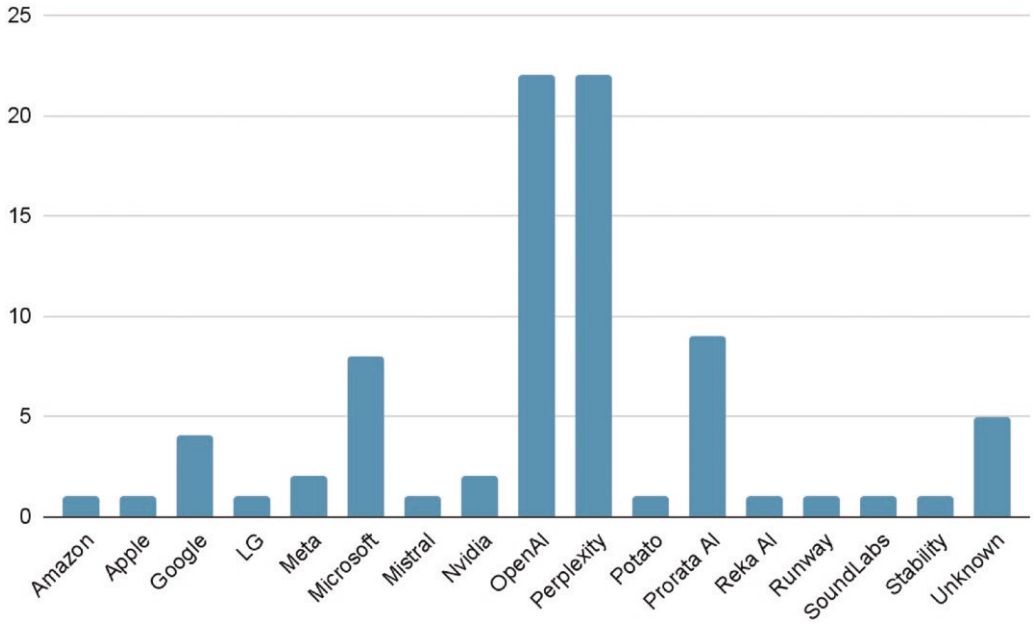

Figure 2 Number of known commercial agreements, by purchaser

Most known commercial agreements are vaguely defined as 'multi-year' deals (23%). Those deals which do specify a time limit (6%) range between two to six years in duration. It is also possible that for agreements where the duration is not specified, that a back-catalogue of content has been purchased, rather than constituting an ongoing year-to-year relationship.

Figure 2 shows that OpenAI and Perplexity have completed the most known commercial agreements with content providers. By contrast, 'Big Tech' companies, including Amazon, Apple, Google, Meta, and Microsoft, have completed far fewer. As noted above, this may be because training data has already been 'licensed' through existing services such that additional licensing agreements are unnecessary. Users of Facebook, for example, agree to let Meta use their public posts and available data for AI training purposes;[26] users of YouTube likewise agree to let Google use their public videos and available data when they use this service.[27]

Figure 3 shows that a number of independent and conglomerate companies have concluded commercial agreements with AI developers. Some companies – notably Shutterstock – have concluded agreements with multiple AI developers. Of the 83 known commercial agreements, at least 12 content providers (14%) are based in the UK. 16% of those agreements which reveal the terms of the agreement are non-exclusive, allowing content providers to contract with other AI developers. Whilst the Reddit/Google agreement was rumoured to be exclusive, subsequent licence agreements with OpenAI suggests that at least some aspects of Reddit content were licensed under a non-exclusive agreement.

Figure 4 shows that most known commercial agreements have been concluded in respect of news and news media content (68%). This is in line with expectations that AI developers rely on content which is both up-to-date and regularly published for the accuracy of outputs. Some other sectors – notably music – make up comparatively smaller percentages, despite their prevalence in policy discussions. It is also possible that types of content are 'bundled' into the sectors presented here – News/Media content may include audiovisual content (and not just textual/informational content).

As noted above, the specifics of the terms of known commercial agreements are limited and in this respect calculating the revenues exchanged between companies is often speculative. The amounts presented here are based on the publicly available information on these known commercial agreements as an aggregate amount of revenue paid to the content provider from the AI developer.[28] Notably, some of this revenue may be offered 'in-kind', with pertinent agreements offering e.g., training opportunities for content creators.

Of the publicly declared revenues from these agreements, we estimate that the news/media sector has secured around USD 285 million; image banks USD 212 million; scholarly/publishing USD 108 million and social media USD 60 million. These figures are conservative estimates raising the possibility that per-annum, or per-usage tariffs will inflate revenues returning to content providers over time.

Indeed, those known commercial agreements which specify the mode of revenue sharing are mostly paid through a percentage amount of usage per item (31% of agreements). Notably, these are the standard terms offered by Perplexity who, as noted above, is one of two most frequent known commercial agreement licensees. Of known commercial agreements, 13% receive revenue

26 'Privacy Policy' <https://www.facebook.com/privacy/policy/> accessed 28 May 2025.

27 'Your content and third-party training' <https://support.google.com/youtube/answer/15509945?hl=en-GB> accessed 28 May 2025.

28 Revenues may have been agreed based on a one-off fee, revenue shares, or per-item tariffs. These figures do not include 'speculative' gains from known commercial agreements, typically based on increases in company-reported revenues which may be 'bundled' with known commercial agreements (eg, Reddit revenues).

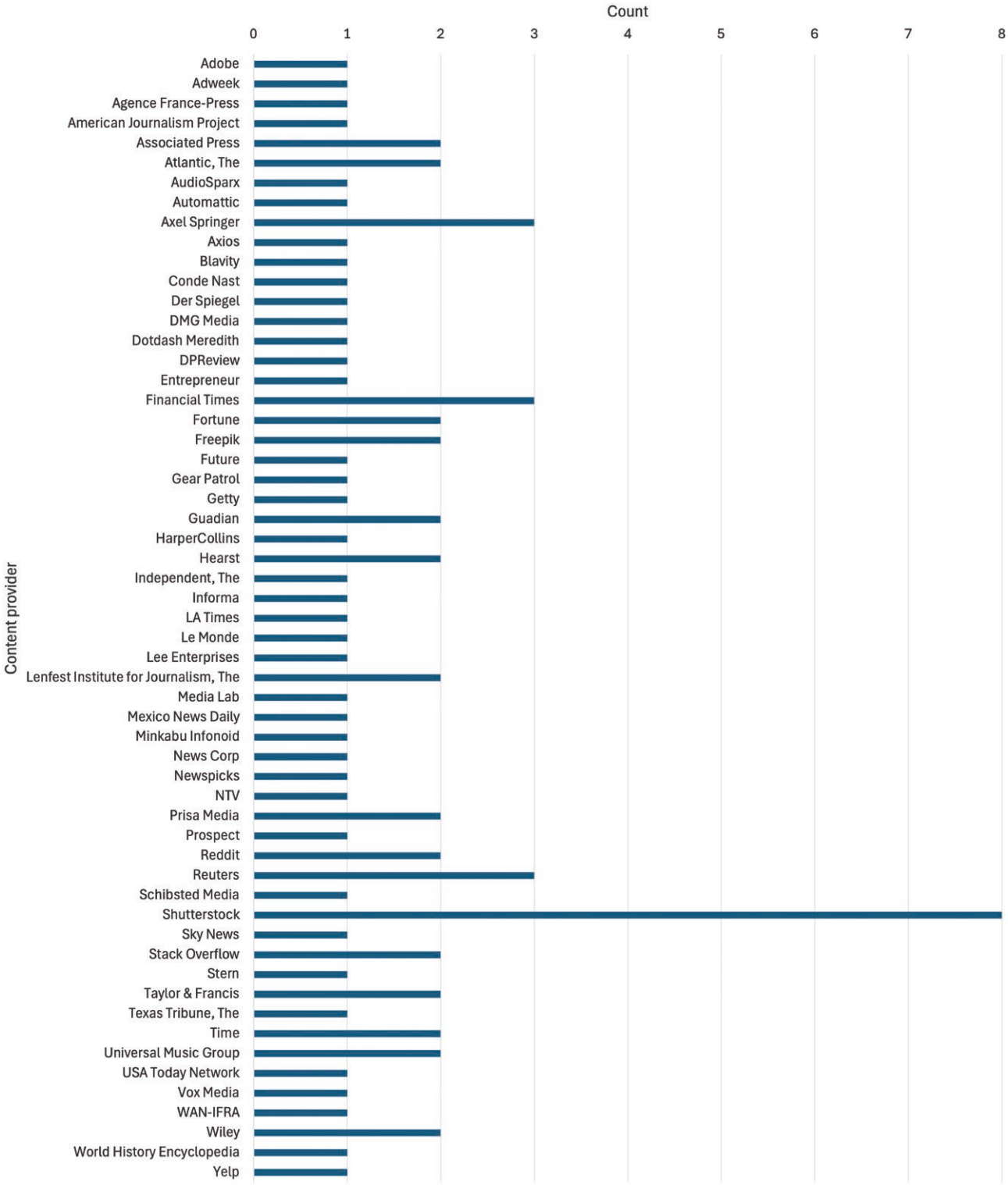

Figure 3 Number of known commercial agreements, by licensor

shares through a percentage of subscription or ad revenue, again forming the standard terms for Prorata.ai, another frequent licensee.

Of note, only eight known commercial agreements specify how payments will be made directly to the primary creators of works being licensed: seven are between Shutterstock and AI developers (payments to be received through a Contributor Fund) and the remainder is HarperCollins/Microsoft (authors receive USD 2,500 if they choose to license their book). While many known commercial agreement announcements specify that methods will be established to ensure this, in fact they lack specific details of how this will be achieved.

2. Income trends for primary creators

There is an ongoing trend of declining revenues for creators in the UK. Figure 5 shows income data from a longitudinal series of surveys on UK creators' earnings.[29]

29 See 'Creators' Earnings Hub' <https://www.create.ac.uk/internal/2024/06/16/creators-hub/> accessed 28 May 2025.

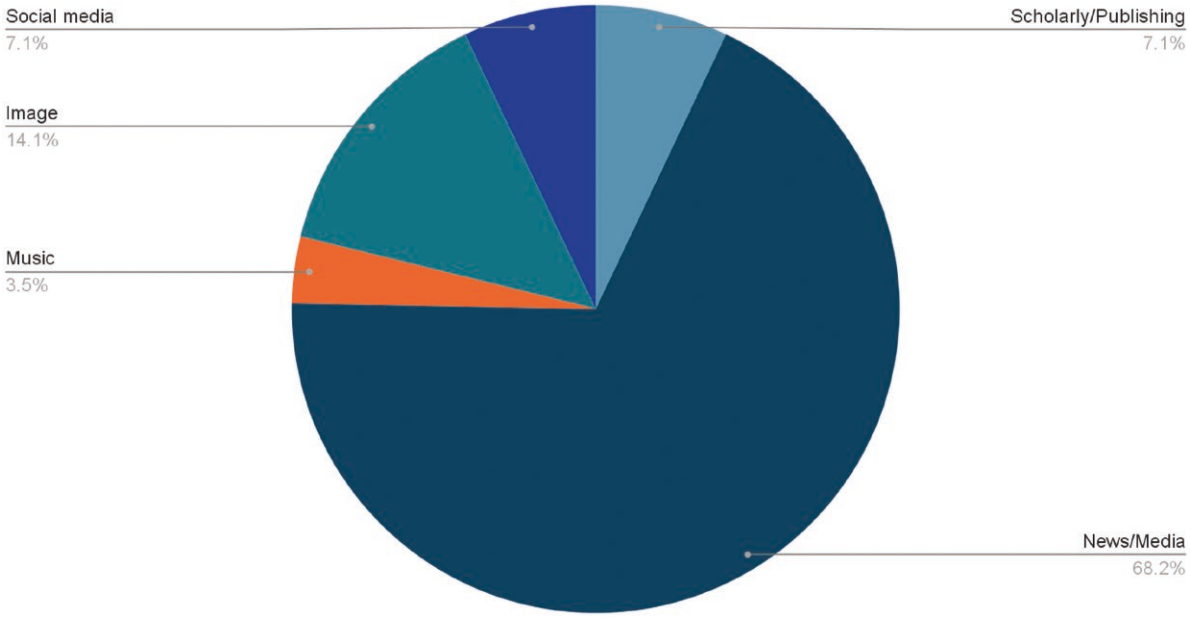

Figure 4 Percentage of known commercial agreements, by sector

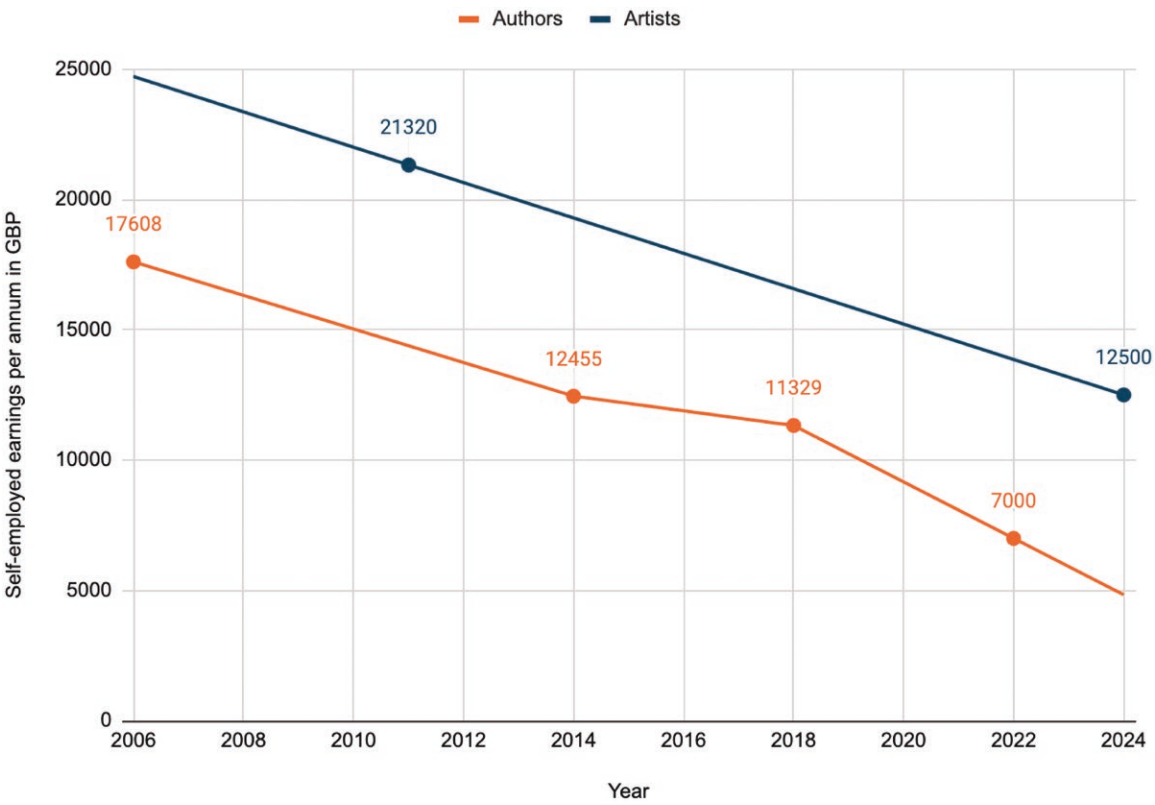

Figure 5 Self-employed median income[30] of creators in GBP per annum

In 2022, authors earned just GBP 7,000 from their creative work – a reduction of 60% since 2006 (GBP 17,608); in 2024, visual artists earned GBP 12,500 per annum – a reduction of 47% since 2010 (GBP 21,320). We consistently find that income from creative activities alone cannot sustain a livelihood consistent with a minimum wage. The reasons for declining earnings are complex and multifactorial. The data indicates that the decline in creators' earnings started before the widespread availability of AI. However, regulating copyright in the AI context offers an opportunity for a wider rebalancing of copyright contracts.

30 This figure excludes income from other sources, including household income (eg, from a spouse) and other individual income (eg, from a second job). It is the income solely earned from a creator's creative work. More information on the methodology adopted here can be found in our latest earnings report: Arthur Ehlinger and others, 'UK Screen Directors: A survey of earnings and contracts' (*zenodo*, 11 February 2025) <https://zenodo.org/records/14689423> accessed 28 May 2025.

3. Open source and local computing

While large commercial AI service platforms have risen to prominence, the proliferation of (partially) open-source AI models, such as Llama 3.3 and DeepSeek R1, is having disruptive effects. The recent trend has been for models to increase in performance while decreasing in storage and compute requirements. Some capable models of fewer than 3 billion parameters can run locally on basic consumer hardware.[31] Intermediary platforms, such as Hugging Face and Ollama, offer access to models as well as software frameworks to facilitate deployment (including commercial use under some licences).

The ability of users to access and deploy AI systems in local computing environments introduces new challenges for regulation and enforcement. On the intermediary side (Ollama, Hugging Face) there is currently no clear liability regime for harms resulting from models made available through these services. While a de facto notice-and-takedown regime appears to be taking shape, the allocation of costs and responsibilities remain ambiguous, and incentives are not adequately addressed. For example, if a user of a large language model causes harm, does liability extend to the intermediary marketplace, to the model developers, or to a service provider?[32] Because models consist of vector embeddings and weights, the results may be unpredictable, making it costly or impossible to ascertain specific risks before a model is used in a harmful way.

As the cost of training models is reduced and the capability of hardware available to consumers is increased, enforcement may become even more difficult. Users are already locally training their own AI models on datasets obtained privately (whether lawfully or unlawfully accessed). It is not clear that any opt-out mechanism would deter such use. Rather, the observed technological trajectory indicates a trend towards widespread bespoke use of AI.[33]

Scientific research and innovation would benefit from clarity about specific activities included in any proposed TDM exception. For example, a typical research project involving AI training will undertake multiple iterative training steps, revisions and fine tuning, potentially spread across multiple sites that might include local university servers or commercial cloud services like Amazon Web Services (AWS) or others. Training data may be obtained via web scraping or through partnership agreements with other institutions or commercial data holders. Data will likely be pre-processed to facilitate training, which could involve format shifting, automatic summarisation or other techniques, before being used in training. After evaluation and benchmarking, models may be re-trained on the same dataset or may be fine-tuned with new sources of data. It is important to note that those models may then become the basis for research and training of other models, underlining the acute need for a permissive legal environment for research. Early-stage R&D could be unfavourably treated under the law compared with incumbent commercial AI services that already benefit from access to massive quantities of licensed data (e.g. commercial publishers and web platforms).

III. Legal analysis

Having reviewed the empirical context of the AI licensing environment, we now address the overall policy direction (presented in the UK government's consultation through option questions 1 and 2), followed by a discussion of six key issues: opt-outs, the exception for data mining, transparency requirements, remuneration of creatives, terms of use, and image rights. Our legal analysis shows that the choices as presented in the consultation are sub-optimal. In our view, rights reservation or opt-out is technically difficult to implement, will increase costs, and create barriers to market entry. The TDM exception (Sec. 29A CDPA 1988) should be broadened to all research before market entry, at which point transparency and licensing obligations would play a role. At the same time, an equitable remuneration provision could be introduced that enables creatives (authors, artists, performers) to receive a share of licensing revenues negotiated by intermediaries (publishers, producers) with AI developers.

1. Opt-out

The UK government's preferred Option 3 is to mirror the EU approach, proposing a wider TDM exception combined with the levers of rights reservation (opt-outs) and transparency requirements. Opt-outs are designed to increase the data available for training, while transparency is supposed to facilitate licensing. The UK government approved explicitly a policy direction to be 'at least as competitive as the EU' for AI innovation.[34]

The negotiations regarding the EU general-purpose AI Code of Practice and early court decisions (*LAION* – Hamburg District Court; *DPG Media* – Amsterdam District Court) have indicated that it will be difficult to operationalise the opt-out. Reasons include: (a) non-standardised or unsuitable technologies (such as robots.txt); (b) modalities (how to opt-out); (c) timing (is it possible after mining has occurred?); (d) location (at the site or at the work level? what about downstream uses?); and (e) complex jurisdictional questions (e.g. if the opt-out takes place abroad).

Difficulties related to the overlap with data protection rights are illustrated in the recent change of Meta's privacy policy. In May 2024, Meta emailed Instagram and Facebook users in the UK to let them know they 'will now rely on the legal basis called legitimate interests for using

31 'Meta unveils slimmed-down Llama models for low-powered devices' (*Silicon Angle*, 24 October 2024) <https://siliconangle.com/2024/10/24/meta-debuts-slimmed-llama-models-low-powered-devices/> accessed 28 May 2025.

32 See Robert Gorwa and Michael Veale, 'Moderating Model Marketplaces: Platform Governance Puzzles for AI Intermediaries' (2024) 16(2) Law, Innovation and Technology 341-91.

33 See Kristofer Erickson, 'IP and the Generative AI Multiverse' (2024) Gikii Conference paper <https://www.create.ac.uk/blog/2023/09/21/ip-and-the-generative-ai-multiverse/> accessed 28 May 2025.

34 Recommendation 24 of the Matt Clifford report.

your information to develop and improve AI at Meta', with the changes due to come into effect on 26 June 2024. Similar messages with key differences were sent to Meta users around the world.

One of these differences was that the right to object, i.e. to opt out, apparently was offered only to UK and EU users under General Data Protection Regulation (GDPR) obligations. The change of policy caused an immediate reaction by digital rights groups,[35] artists organisations,[36] and data protection authorities and regulators,[37] with the availability and complexity of the opt-out process at the centre of the complaints.

This experience raises overarching questions which affect the main areas of the consultation, including the opportunity, workability and effectiveness of mechanisms for allowing rightsholders to reserve their rights. Effective modalities for such reservations, enforcement across different jurisdictions where online platforms adopt responsive policies, and whether opt-outs have any effect on the actual training of AI models by big tech companies remain complex challenges.

The uninformed nature of consent to use users' data and content in AI training is another fundamental issue. It is interesting to note that Meta did not feel compelled to change their copyright policy together with that on privacy, perhaps relying on the non-exclusive licence to which users (mostly unknowingly) agree when they create a Facebook or Instagram account.[38] For different reasons and in different ways, copyright law and data protection law pose similar challenges for companies, researchers and institutions developing and using AI technology.

2. Text and data mining exception

The existing TDM exception in UK law (Sec. 29A CDPA) is not fit for purpose and does not reflect current research practice. In our response to the UK Intellectual Property Office (IPO) call for views on Artificial Intelligence and IP: copyright and patents (2021-2022),[39] we encouraged the government to consider two options not included in the consultation: (i) excluding from the scope of exclusive rights TDM and other acts of extracting informational value from protected works;[40] or (ii) introducing a technologically neutral, open-ended exception akin to the fair use doctrine in the USA.

Statutorily excluding AI training and other technical uses from the scope of copyright (i)[41] or leaving this determination to courts on a case-by-case basis (ii) remain plausible options, which are adopted in a variety of jurisdictions with strong AI economies.[42] In light of the developments that followed the previous consultation,[43] we propose a third option for consideration, which may be more realistic in the current UK policy environment: a broadened Sec. 29A that covers all research and development before market entry, at which point transparency and licensing obligations apply.

As the AI Opportunities Action Plan suggests, EU law is a key reference point for reform. Article 3 of the Copyright in the Digital Single Market Directive (2019/790, the CDSM Directive) provides a mandatory exception to copyright and database rights to carry out text and data mining for the purposes of scientific research. While under EU law the purpose itself does not have to be 'non-commercial', the exception is only available to research organisations (not-for-profit) and cultural heritage institutions.

On the other hand, the UK TDM exception offers the advantage of being available to everyone but can only be relied upon for research for non-commercial purposes. This restriction hinders the functioning of the UK research environment (as explained in the evidence discussion above). It was inherited from another EU Directive, namely the Information Society Directive (2001/29/EC).[44]

35 Complaints to Data Protection Authorities (DPAs) from None of Your Business (NOYB) in the EU led Meta to pause and review their plans to use user data to train generative AI <https://www.dataprotection.ie/en/news-media/latest-news/dpcs-engagement-meta-ai> accessed 28 May 2025. However, as no official change was made to Meta privacy policy to make this legally binding, Open Rights Group (ORG) raised a formal regulatory complaint in the UK. The full complaint and the initial response of the Information Commissioner Office (ICO) can be found here <https://www.openrightsgroup.org/press-releases/org-complaint-to-ico-about-meta-privacy-policy-changes/> accessed 28 May 2025. Later in 2024, the ICO closed down the ORG's complaints stating that they were satisfied by the changes made by Meta to the opt-out process, including making it simpler and available for a longer window. ORG argues that the opt-out process remains unnecessarily difficult and that opt in consent should apply to Meta <https://www.openrightsgroup.org/press-releases/icos-failure-to-enforce-is-putting-the-public-at-risk/> accessed 28 May 2025.

36 See for example DACS, 'What artists and their beneficiaries need to know about Meta's new privacy policy' (*DACS*, 30 May 2024) <https://www.dacs.org.uk/news-events/what-artists-and-their-beneficiaries-need-to-know-about-metas-new-privacy-policy> accessed 28 May 2025.

37 In July 2024, Brazil's National Data Protection Authority (ANPD) suspended with immediate effect the validity of Meta's new privacy policy <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta> accessed 28 May 2025.

38 See for example Instagram's Terms of Use, 4.3 Permissions you give to us <https://help.instagram.com/termsofuse> accessed 28 May 2025.

39 Martin Kretschmer, Bartolomeo Meletti and Luis H Porangaba, 'Artificial intelligence and intellectual property: copyright and patents – a response by the CREATe Centre to the UK Intellectual Property Office's open consultation' (2022) 17(3) JIPLP 321-26 <https://academic.oup.com/jiplp/article/17/3/321/6550465> accessed 28 May 2025.

40 An example of this approach is art 30-4 of the Japanese Copyright Act, which explicitly states that it is permissible to exploit a work when the purpose of the use is not 'to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work'. Chosakukenhō [Japanese Copyright Act] 1970, art 30-4 <https://www.wipo.int/wipolex/en/text/584871> accessed 28 May 2025.

41 In UK copyright law, for example, this could be implemented by broadening s 28A CDPA (Making of temporary copies).

42 For a recent overview, see Matthew Sag and Peter K Yu 'The Globalization of Copyright Exceptions for AI Training' (2025) 74 Emory Law Journal (forthcoming) <https://ssrn.com/abstract=4976393> accessed 28 May 2025.

43 The main outcome of the previous consultation was to introduce a new TDM exception for any purpose. Following the reaction of the creative industries and a recommendation by the House of Lords Communications and Digital Committee, on 1 February 2023 the Minister for Science, Research and Innovation stated in parliament that the proposal to introduce a new TDM exception would not go ahead. In June 2023, the IPO set up a working group with representatives from the technology, creative and research sectors to agree on a code of practice on copyright and AI. On 6 February 2024, the government announced that the working group failed to agree on an effective voluntary code <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 28 May 2025.

44 The InfoSoc Directive imposed the same non-commercial limitation also to s 29 (Research and private study) and s 32 (Illustration for instruction) CDPA.

The UK now has the opportunity to devise a new approach to copyright and research in the age of AI. An approach which is forward-looking and in line with the common law tradition, enabling research excellence to flourish for the advancement of science and technology, while also encouraging the curation of high-quality data, responsive licensing and fair remuneration. To achieve these ambitions, we propose the following amendments to Sec. 29A:

- The 'non-commercial' purpose restriction should be removed. It creates uncertainty, prevents public-private partnerships, and conflicts with the UK legal tradition. The scope of the exception should be limited by market entry.
- The exception should remain available to everyone, and focus on a broader research purpose that can capture how R&D will evolve in the future.
- Sharing and retaining copies made under the exception should be allowed to the extent justified by the permitted purpose. This is to reflect current research practice, including knowledge transfer, public-private partnerships, and experimentation.[45]
- The exception should remain non-overridable by contract [Sec. 29A(5)].
- The exception should apply also to the sui generis database right to reduce uncertainties around its application.
- A more streamlined mechanism for the exception's beneficiaries to remove technological protection measures (TPMs) is needed to unlock data across the UK research infrastructure.[46]

The proposed approach comes with challenges; the main ones relating to how to determine the market entry point, and what happens when a product enters the market. There are several legislative routes that can be explored to determine the point of market entry, such as focussing on selling and letting for hire,[47] the notion of experimental use in patent law[48] and design rights,[49] defining the meaning of 'research and development' through statute and guidelines,[50] or leaving the interpretation of notions such as 'R&D' and 'market entry' to courts. It will require careful consideration, but it is possible.

An extended TDM exception should not affect licensing deals with content providers. Beneficiaries of the expanded exception would still be willing to purchase access to high-quality content. In fact, this approach may lead to new potential income streams and thus positive incentives for content providers. They would be willing to offer curated datasets whose use is easy to document and then track once a product developed under the exception enters the market.

A similar positive incentive may be created on the side of the beneficiary with limited transparency obligations in R&D. If copyright-related transparency obligations apply only at the point of market entry, purely experimental projects with no commercial horizon would be free from bureaucracy, whereas those with commercial ambitions would be motivated to maintain proper records. While most of the dealing in the market would be left to the market, some mechanisms to enable the commercial exploitation of products developed under the exception may be needed. At least conceptually, Extended Collective Licensing (ECL)[51] could have positive horizontal effects across the copyright system, such as facilitating mass digitisation efforts and the equitable remuneration schemes proposed below.

It is important to note that a truly open research infrastructure, where researchers across disciplines have ready access to data and can contribute to the infrastructure's development with their own research data, requires structural solutions. Any legislative intervention in this space will have to strike a balance between leaving the language of the law open enough to capture future changes in technology and practice, and give beneficiaries sufficient detail and clarity to make informed decisions.

An important component of a functioning research infrastructure are mechanisms to continuously identify the uncertainties and needs faced by researchers across different disciplines and sectors, and to develop guidance

45 In its current form, s 29A(2)(a) stating that copyright is infringed when a copy made under the exception is 'transferred to any other person' is too restrictive. Combined with a conservative interpretation of the 'lawful access' requirement, it poses major obstacles in research partnerships, such as the Living with Machines example mentioned above.

46 The cumbersome user complaint process provided in s 296ZE CDPA to remove TPMs when they interfere with permitted acts was also inherited from the InfoSoc Directive and is being used rarely. Over a period of ten years, the UK IPO has received only three complaints. See Anthony Rosborough, 'Technological Protection Measures & the Law: Impacts on Research, Education & Preservation' (*zenodo*, 25 November 2024) <https://doi.org/10.5281/zenodo.14172278> accessed 28 May 2025. Another interesting reference for the UK government in this area is the German implementation of art 17 CDSM, in particular art 18(4) Act on the Copyright Liability of Online Content Sharing Service Providers. It provides that after an abusive blocking request for works in the public domain or whose use is permitted by law to anyone free of charge (ie under an exception or open licence), the service provider (eg YouTube) must ensure that these works are not blocked again.

47 The existing TDM exception already treats as infringing copies those made under the exception and then 'subsequently dealt with' [s 29A(3)], defining 'dealt with' as 'sold or let for hire, or offered or exposed for let or hire' [s 29A(4)]. A similar device is used, inconsistently and at times problematically, across the CDPA, including in s 31A(6) and (7), s 31B(11) and (12), s 35(5) and (6), s 36(8) and (9). In s 35 ('Recordings by educational establishments of broadcasts', the exception underpinning the Educational Recording Agency licence) and s 36 (Copying and use of extracts of works extracts of works by educational establishments', the exception underpinning the Copyright Licensing Agency licence), in addition to selling and letting for hire, 'dealt with' also means 'communicated otherwise than as permitted by subsection (2)'. In both sections, subsection (2) allows the communication of copies made under the exception by or on behalf of the educational establishment to its pupils or staff for 'the non-commercial educational purposes of that establishment' [s 35(2)] or for 'the purposes of instruction for a non-commercial purpose' [s 36(2)].

48 s 60(5)B UK Patents Act 1977.

49 s 244A(b) CDPA.

50 For example, s 1138 of the UK Corporation Tax Act 2010 defines 'research and development' as 'activities that fall to be treated as research and development in accordance with generally accepted accounting practice.' The HMRC guidelines on the 'Meaning of research and development for tax purposes' (March 2023) define R&D as including 'when a project seeks to achieve an advance in science or technology', 'activities which directly contribute to achieving this advance in science or technology through the resolution of scientific or technological uncertainty', and 'certain qualifying indirect activities related to the project.' <https://www.gov.uk/government/publications/guidelines-on-the-meaning-of-research-and-development-for-tax-purposes/meaning-of-research-and-development-for-tax-purposes-guidelines> accessed 28 May 2025.

51 'Extended Collective Licensing (ECL): Guidance for relevant licensing bodies applying to run ECL schemes' <https://www.gov.uk/government/publications/extended-collective-licensing/extended-collective-licensing-ecl-guidance-for-relevant-licensing-bodies-applying-to-run-ecl-schemes#fn:2> accessed 28 May 2025.

and best practices that respond to those needs.[52] Issues such as those faced by the Living with Machines project (see above) should be addressed at an infrastructural level, not on a project-by-project basis. These are complex issues that require a collective risk management approach.

3. Transparency

The need for transparency requirements is one of the less controversial aspects of AI governance, with numerous regulators around the world proposing their adoption.[53] Most of these proposed obligations apply to generative AI and copyright-protected works. We agree that AI developers should disclose the sources of their training material. Among other benefits, it can facilitate the identification of copyright protected content and potential licensing.

We identify three high-level approaches that the UK could follow to increase AI transparency, which come with both shared as well as their own specific challenges:

a) Voluntary and industry-led transparency initiatives

The UK IPO has experience with this form of governance arrangement, having previously operationalised and worked as a facilitator in the development of industry-led initiatives, such as the UK Voluntary Code of Good Practice on Transparency in Music Streaming.[54] However, the failure of the UK IPO-led working group to reach an agreement on a voluntary code of practice on AI and copyright shows the limitations of this regulatory approach in contested fields. More generally, meaningful AI transparency may be difficult to achieve based on voluntary commitment only.[55]

b) Horizontal or specific transparency obligations

In the US, the California Bill (Assembly Bill No. 2013) provides a non-exhaustive list of the data used to train generative AI systems or services (a high-level summary of the datasets) that should be made publicly available by AI developers, including any data protected by IP rights or in the public domain. The main purpose is to enable consumers to evaluate and compare the AI systems they use and to take action to address any shortcomings in those systems.[56] These transparency obligations can be privately enforced, but there is no public oversight or enforcement mechanisms to ensure compliance.

In parallel, there is also a proposal for a Generative AI Disclosure Copyright Act (H.R.793) at the federal level, which provides for specific transparency obligations under copyright law. These include requiring any person who creates or significantly modifies a training dataset to file a notice with the Register of Copyrights that provides for a sufficiently detailed summary of any copyrighted works used for these purposes and the URL for those datasets.[57] The register is responsible for creating a publicly available database that contains all the notices. Although a civil penalty can be imposed for non-compliance, the responsibility for monitoring and filing a complaint rests with the copyright owners.

There is a vast literature and evidence about the limitations of these types of transparency requirements in different fields, especially those that rely on individuals' or consumers' action for compliance. As scholars have noted, it is not only the consumer behaviour that should be taken into account, but all stakeholders in the regulatory system.[58] Transparency may not be that effective for consumers and individual creators, but it can be useful for large rightsholders and collecting societies when negotiating licensing agreements with AI developers.

c) Horizontal transparency obligations followed by accountability and enforcement mechanisms

The EU AI Act establishes horizontal transparency obligations for providers of general-purpose AI (GPAI) models in Art. 53. In line with the copyright legislation, Art. 53(1)(d) requires GPAI model developers to draw up and make publicly available a sufficiently detailed summary about their training data, according to a template provided by

52 CREATe has co-developed methodologies aimed at identifying the specific copyright related issues faced by certain communities and enabling them to respond to those issues through the development of best practices. The methodological approach, which builds upon similar initiatives in the US (Codes of Best Practices in Fair Use) as well as deliberation theory and methods, is described in Bartolomeo Meletti and Stef van Gompel, 'Encouraging the Use of Copyright Flexibilities through Codes of Best Practices: A Deliberative Approach' in Caterina Sganga and Tatiana Synodinou, *Flexibilities in Copyright law* (Routledge 2025) forthcoming. The methodology has proved successful in capturing sector-specific issues and encouraging positive change in risk-averse behaviour. For an overview of different best practices projects, see <https://www.create.ac.uk/project/creative-industries/2022/12/07/developing-best-practice-codes-for-creative-audiovisual-re-use/> accessed 28 May 2025.

53 For example, EU AI Act, South Korea's Framework on the Development of Artificial Intelligence and Establishment of Trust Foundation (AI Basic Act), Brazil's Bill No 2,338/2023 (Brazil's Proposed AI Regulation), Canada's Bill C-27, the Digital Charter Implementation Act, 2022 (Canada's Artificial Intelligence and Data Act – AIDA).

54 UK Voluntary Code of Good Practice on Transparency in Music Streaming (31 January 2024) <https://www.gov.uk/guidance/uk-voluntary-code-of-good-practice-on-transparency-in-music-streaming> accessed 28 May 2025.

55 In the US, during the Biden administration, leading US AI companies signed voluntary commitments with the White House to develop AI in a safe and trustworthy way, including to increase transparency. However, while progress has been made in this area, there has been no meaningful transparency or accountability from these AI companies. Despite the limitations of these governance mechanisms, the Trump administration seems to be following a similar path, even more flexible, prioritising public-private partnerships to set standards and risk levels and operationalise codes of conduct, as his recent executive order on AI suggests. See Melissa Heikkilä, 'AI Companies Promised to Self-Regulate One Year Ago. What's Changed? (*MIT Technology Review*, 22 July 2024) <https://www.technologyreview.com/2024/07/22/1095193/ai-companies-promised-the-white-house-to-self-regulate-one-year-ago-whats-changed/> accessed 17 February 2025; 'Removing Barriers to American Leadership in Artificial Intelligence – Executive Order' (23 January 2025) <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/> accessed 28 May 2025.

56 <https://leginfo.legislature.ca.gov/faces/billAnalysisClient.xhtml?bill_id=202320240AB2013> accessed 28 May 2025.

57 H.R.7913 – Generative AI Copyright Disclosure Act of 2024 <https://www.congress.gov/bill/118th-congress/house-bill/7913/text> accessed 28 May 2025.

58 Julia Black and Andrew Murray, 'Regulating AI and Machine Learning: Setting the Regulatory Agenda' (2019) 10 European Journal of Law and Technology 201.

the AI Office. However, different from the California Bill, the EU Act provides for a governance structure and framework to ensure compliance.

The AI Office plays a key role in implementing the AI Act and enforcing the rules for the GPAI models. Failure to comply with these transparency obligations may result in administrative fines (up to 3% of their annual total worldwide turnover or EUR 15,000,000, whichever is higher), as established in Art. 101. South Korea has adopted a similar approach, with horizontal transparency obligations and a comprehensive governance framework that includes the Ministry of Science and ICT (MSIT), the National AI Committee and the AI Policy Centre.[59] Under the Korean AI Basic Act, which is expected to come into effect in 2026, the MSIT has oversight and enforcement powers.

In our view, voluntary initiatives are likely not sufficient. Some statutory underpinning may be necessary to ensure compliance with transparency requirements, whether followed or not by public oversight and enforcement mechanisms. To make these policy choices, it is necessary to define transparency obligations more precisely.

d) When and to whom should the transparency obligations apply?

In order to limit bureaucracy and costs, transparency obligations for AI developers should only apply at the point of market entry of AI models. Thus, AI developers who plan to place or put into service their generative AI models or systems in the UK market would be required to keep up-to-date records of the data, including copyright-protected content, used to train their models. At the same time, this would reduce the legal risks for AI developers, especially of copyright infringement, in the research and development (R&D) stage.

Both the EU AI Act[60] and California Bill[61] adopt a similar scope. Their transparency rules apply to 'providers' or 'developers' whose models and/or systems are put in place, put into service or made available on the market for use by the public. In California, this includes developers who substantially modify an AI system or service. However, in both laws, web crawlers are not explicitly covered, although they may carry out copyright-relevant acts. Similar to the EU AI Act, the specific needs of small businesses and new entrants must be considered when adopting transparency obligations to ensure that requirements are not disproportionately burdensome and costly for them.[62]

e) What transparency obligations should they have?

To ensure meaningful transparency and encourage compliance, the law needs to be clear about what kind of transparency is required.[63] This is why recent transparency legislation, such as the EU Digital Services Act (DSA) and the UK Online Safety Act (OSA), provide detailed information on what data should be disclosed and how. They also specify when further action (e.g., EU Commission delegated acts) including by private actors (e.g., codes of practices) is required.

In our view, some form of agreement between industry and government on the level of granularity is necessary. The EU AI Office's template for the summary about the training data and accompanying guidance, developed under Art. 53(1)(d) of the AI Act, provides a good starting point for defining the type of information and the minimum level of detail that AI developers should cover for compliance. For example, the proposed template suggests the categories of information sources and the characteristics/information about the data sources that should be presented in the summary. The level of detail that should be included in the law is one of the challenges in ensuring meaningful transparency as well as future-proofing legislation.

Similar to the California Bill in Sec. 3111, a non-exhaustive list of information, such as 'whether the datasets include any data protected by copyright, trademark, or patent, or whether the datasets are entirely in the public domain' and 'whether the datasets were purchased or licensed by the developer', could be included in UK law. This could be followed by a statutory mandate for a government body, such as the UK IPO, to work with key stakeholders to develop a template or code of practice specifying the level of detail and how the data should be disclosed. By establishing mandatory transparency obligations combined with this statutory mandate, the UK government could ensure that some form of agreement can be reached, which was not possible through voluntary mechanisms.

f) How should these transparency obligations be implemented to encourage compliance?

In our view, transparency obligations regarding training data should be horizontal and cross-sectoral, as the use of data to train generative AI models involves more than just copyright issues (see below). In principle, a tiered approach to transparency could also ensure that relevant actors receive the information they need and that data protected by trade secrets and confidential commercial information does not become an obstacle to transparency. For example, the information copyright owners have the right to access or request may be different from the information AI developers should make publicly available. In this case, a public authority could be designated to mediate this access to information, rather than relying on the courts.

59 Kim Eun-Jin, 'AI Basic Act Passes National Assembly, Aiming for Enhanced AI Reliability and Regulation' (*Business Korea*, 27 December 2024) <https://www.businesskorea.co.kr/news/articleView.html?idxno=232661> accessed 18 February 2025.

60 Under art 53 of the AI Act, the EU applies its transparency obligations only to 'providers' that develop a GPAI model or have a GPAI developed and place it on the market or put the GPAI system into service under their own name or trademark, whether for payment or free of charge'; art 3(3) of the AI Act defines GPAI providers.

61 California, in s 3111 of Assembly Bill n 203, establishes that the obligations should fall on developers who design, code, produce, or substantially modify an AI system or service for use by members of the public; s 3110(b) of the California Bill defines generative AI developers.

62 In the EU, the Second Draft of the Code of Practice sets out separate Key Performance Indicators for Small and Medium-size Enterprises (SMEs) <https://artificialintelligenceact.eu/small-businesses-guide-to-the-ai-act/> accessed 28 May 2025.

63 Chris Reed, 'How should we regulate artificial intelligence?' (*the Royal Society*, 6 August 2018) <https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0360> accessed 28 May 2025.

g) Extraterritoriality and international interoperability

In the EU, the transparency and compliance obligations under Art. 53 of the AI Act imposed on the GPAI model providers are presented as having an extraterritorial effect, and could apply to models trained outside the EU.[64] From a private international law perspective, there are many questions about the legitimacy of the extraterritoriality of EU law. In designing its legislation, the UK will also need to decide whether its transparency obligations should apply to all AI models placed on the UK market, including in relation to the entire AI lifecycle, regardless of whether they were developed and trained in the UK. AI companies will inevitably face global challenges when exporting AI. There is also a need for international cooperation to ensure the international interoperability of information and the development of transparency standards.

h) Open source models

Open-source AI models are a particular case that warrants further investigation. First, they present significant challenges in determining which entities should undertake transparency obligations. This issue becomes particularly complex when open-source AI developers are based outside the UK, while deployers or downstream developers utilise these models within the UK market. In such cases, the key question remains: who should bear responsibility for ensuring compliance with transparency obligations? On one hand, downstream users often lack sufficient knowledge about the training dataset, the specific methodologies used to train the AI, and potential copyright issues. Even when fine-tuning, users' understanding would be limited to the modifications they introduce, rather than the full development history of the model. On the other hand, directly imposing transparency obligations on open-source AI developers located outside the UK poses significant jurisdictional and enforcement challenges. Extending transparency obligations extraterritorially would most likely prove impractical, as enforcing compliance across jurisdictions presents hurdles. Meanwhile, because of the open-source nature, it would disrupt the understanding of what constitutes 'entry into the UK market,' as it would require defining whether merely publishing an open-source AI model online – without actively offering it as a service – qualifies as market entry.

4. Remuneration

As noted in the Evidence review (Section II above), many holders of copyright (or exclusive licensees), such as publishers, have entered into arrangements with AI developers to provide them with access to their datasets. In some circumstances, rightsholders have sought the author's permission. As far as we are aware, only occasionally (as with Shutterstock and HarperCollins) has any provision been made by publishers/intermediaries to remunerate authors or distribute a share of the proceeds of the arrangements with AI producers.

In an environment in which creators are concerned about the impact of AI on the continued marketability of original works, it seems appropriate that authors should be entitled to share in the proceeds of these deals. Indeed, at the time that intermediaries entered the original contractual relations with creators, AI uses were not likely envisaged.

We propose imposing a duty on licensors of data sets for AI development to share the proceeds of such deals with the authors and performers whose works are contained in the dataset. Two precedents for such rights already exist in UK law: the right to equitable remuneration where the rental right is transferred (CDPA, Sec. 93B) and the right of performers to claim equitable remuneration when sound recordings are played or communicated to the public (CDPA, Sec. 182D).

As with those models, we would propose the amount of equitable remuneration be agreed by the parties (the creators and the provider of the dataset), but in default of agreement to be determined by the Copyright Tribunal. The parties are likely in a better position to determine equitable division of the remunerations, as for example, between recently published and material that has been accessible for decades. It seems likely that competition from AI-produced material is more likely to damage expected revenues from newly created than older material. We would also propose that the right not be assignable except to a collective management organisation (which might operate on behalf of groups of creators). Finally, we would suggest that the right to equitable remuneration cannot be compromised by contract. A model for this might be Sec. 191HB(7) ('An agreement is of no effect in so far as it purports to exclude or restrict the entitlement').

We have already noted that the terms of such deals are rarely made public but rather treated as a matter of commercial confidentiality. In order to render the creators' entitlements effective, we would suggest supplementing the duty to pay such remuneration with an ancillary transparency duty on intermediaries to reveal these details, albeit in confidence, to the authors whose works were included in the datasets. EU law recently recognised such a right in Art. 19 of the Directive on Copyright in the Digital Single Market.

We recognise that the arrangements entered by publishers and intermediaries with AI developers are not entirely analogous to the existing remuneration rights. In particular, those rights are parasitic upon particular exclusive rights (rental, communication to the public) which have been transferred or allocated to the intermediary. The scope of the right thus controls the scope of the entitlement to remuneration. As explained in Section III(2), we do not propose that an explicit right to train data is recognised. At least until the launch of AI products, such training should fall within the text and data mining exception. However, insofar as the parties to these exploitation agreements appear to believe that the reproduction right is implicated, and the arrangements are being made on that premise, it seems right that authors are entitled to a share of these revenues.

64 Séverine Dusollier and others, 'Copyright and Generative AI: Opinion of the European Copyright Society' (2025) 16 JIPITEC 121 <https://www.jipitec.eu/jipitec/article/view/424> accessed 28 May 2025.

5. Terms of use

Paragraph 92 of the Consultation raises a question about the standard Terms and Conditions (T&Cs) of generative AI models. It focuses on the validity and potential power imbalance in the assignment of licences for the work of creators, performers, and other right holders. Indeed, a recurring and important aspect with regard to the copyright provisions in the T&Cs of generative AI models was that the ownership of outputs was assigned to the user, but, in many cases, an extensive licence was also granted back to the model for co-existing use of the outputs.[65]

This is reflected indirectly in the UK Copyright and AI Consultation which, when discussing aspects of contracts and licensing for creators, states that 'standard terms and conditions for some services require works to be licensed on a broad basis which provide little choice to their users'.[66] The scenario seems to be becoming more problematic, with the emergence of a new trend where the grant of output ownership to users is conditional upon certain requirements. Often such a requirement is a paid subscription to the service providing the GenAI model.[67]

This new way of granting output ownership may result in a more fragmented self-regulation (private ordering) of generative AI providers, with a 'switch' of the providers from neutral content hosts ('platformisation' paradigm) to owners of AI-generated outputs, even when these are prompted by users and may not attract copyright in the first place.

Fairness and power imbalances between GenAI companies and users should be framed under wider principles considering the relationship with right holders and users. T&Cs of Generative AI models are a complex challenge, since 'users' of GenAI are a diverse group with diverging interests. Paragraph 92 of the Consultation rightly focuses on rightsholders as users – and the control rightsholders have in their relationship with GenAI providers. However, T&Cs are likely to apply quite broadly, casting the net widely (as acknowledged in para. 92: '[I]t may be quite difficult for right holders to control the use of their work with AI, as terms relating to this may be bundled with other contractual terms or framed in broad or vague language').[68] In this regard, there are existing powers and mandates which allow a redress of a bargaining imbalance in T&Cs, on a more granular specific level, i.e. narrowed down to a subset of users.

These powers rest with the Competition and Markets Authority (CMA) under the UK Digital Markets, Competition and Consumers Act 2024 (DMCC Act). Under the DMCC Act, the CMA can protect users against the exploitative conduct of designated gatekeepers, which could include AI services.[69] The CMA has indicated, in its Invitation to Comment on Strategic Market Status (SMS) investigation into Google's general search and search advertising services, that it could consider interventions targeting unreasonable terms on consumers (particularly in relation to the collection and use of consumer data without informed consent) but also the use of publisher content without fair terms and conditions (including payment terms, an issue addressed in Section II).

These interventions may also be possible in the context of other designations under the DMCC Act, or provide principles for other (regulatory) interventions. Thus, we suggest that there ought to be a holistic approach to T&Cs for the use of rightsholder content across the CMA and other authorities. Moreover, the imbalances in bargaining position may be more pronounced when considering groups of creators who do not possess the collective bargaining strength of news publishers.

In summary, we suggest relying on legal solutions that have been implemented elsewhere and we believe that a requirement for generative AI services to have fair T&Cs can be pursued with the intervention of the CMA implementing the DMCC Act. This would lead to a more holistic approach in preventing harms, such as ownership of outputs and content moderation, faced by all users, including creators.

6. Image rights

At present, the UK does not have a distinct image right, personality right, or right of publicity. Despite the absence of such a right, it is standard practice that media entities and commercial firms seek permission, and make a contract, when aiming to use any celebrity's persona or image in an endorsement or a piece of advertising.[70]

The consultation focused on the possibility that new AI technologies could enable widespread misuse or misappropriation of image/persona via digital replicas. Notably, artists and celebrities fear losing control over their image in the context of such replicas. Below, we consider various aspects of this debate: types of outputs, neighbouring rights, passing off, *sui generis* protection, and data protection. We note that concerns about 'deep fakes', which are used for nefarious purposes such as fraud and misinformation, are more appropriately addressed by criminal law and technological detection and takedown measures, rather than via intellectual property law.[71]

a) Regulation of outputs

The creation of digital AI replicas involves inputs and outputs. It is the position of CREATe that an

65 See Lilian Edwards and others, 'Private Ordering, Generative AI and the "Platformisation Paradigm": What Can We Learn from Comparative Analysis of Models Terms and Conditions?' (2025) 1 Cambridge Forum on AI: Law and Governance e2.

66 'Copyright and Artificial Intelligence' para 92 <https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence> accessed 28 May 2025.

67 See for example 'Terms of Service' (AI Image Generator – Create Art, Images & Video | Leonardo AI) para 8.3 <https://leonardo.ai/terms-of-service/> accessed 28 May 2025.

68 'Copyright and Artificial Intelligence' (n 66) para 92.

69 See CMA, 'AI Strategic Update' (*CMA*, 29 April 2024) <https://www.gov.uk/government/publications/cma-ai-strategic-update/cma-ai-strategic-update> accessed 28 May 2025.

70 Emma Perot, *Commercialising Celebrity Persona* (Hart 2023).

71 Emma Perot and Frederick Mostert, 'Fake It Till You Make It: An Examination of the US and English Approaches to Persona Protection As Applied to Deepfakes on Social Media' (2020) 15 Journal of Intellectual Property Law and Practice 32.

input should benefit from the research-focused training approach advocated throughout this consultation response. As such, the focus of any potential legal intervention should be on the potential IP ramifications of the output.

It should be noted that transformative output will likely fall outside any potential *sui generis* image right or personality-based protection. This is because the combination of elements of persona from multiple sources can mean that the output does not reflect a true 'replica' of any one persona.

b) Neighbouring Rights

In alignment with the 'opt-in' approach, if their performances have been used in training, performers could be remunerated under the existing neighbouring rights in the CDPA 1988, which would potentially trigger when the AI product goes to market. However, neighbouring rights are limited to capturing performances and do not confer the protection of persona in the abstract.[72] Overall, the protection of performers rights is limited in scope and will be subject to contractual arrangements. Therefore, this option offers only limited redress for performers affected by digital replicas.

c) Passing off

It remains an open question as to how a case involving an unauthorised digital replica of a person would be dealt with under the tort of passing off. Passing off is primarily concerned with consumer-based harm. There are three elements: (a) goodwill or reputation in a name, image, likeness etc; (b) misrepresentation to the public, e.g. such that members of the public are likely to believe that the claimant has endorsed the defendant's goods or services; (c) damage/harm.

Can unauthorised commercial use of a digital replica be subject to an action under passing off? It is feasible that the English courts could develop the tort to include digital replicas of famous celebrities. While passing off has traditionally applied to protect against false endorsement in merchandising and advertising, as a common law tort action, passing off could offer enough flexibility to be extended to other fora which use personas, such as movies, television, and video games. However, the law of passing off, as it currently stands, is unlikely to provide an avenue for redress to individuals who do not meet the current standard of goodwill.[73] In other words, only well-known actors/celebrities, i.e. those with a substantial reputation, are likely to be able to avail of the tort. This potentially limits legal recourse for small-scale artists such as background actors and voice actors.

d) Sui generis protection and its potential consequences

It is our view that it is premature to introduce a new statutory right such as a *sui generis* image right, personality-based right, or AI replica right. Evidence from various industries would be needed to understand fully and predict the implications, transaction costs, and normative trade-offs involved. As noted above, the long-standing position in the UK is that image, voice, and likeness are not protected as such. This traditional UK position has enabled a permissive environment, which has set rational standards of behaviour, and created expectations that could be shifted by a new statutory right.

Perhaps of greatest concern is the unintended consequence of limiting freedom of expression. Any new statutory right would have to be tightly defined and limited in *scope* to ensure that biographical and historical information could be conveyed, and so that news reporting would not be unduly hindered.

In addition to scope, considerations on the *term* of protection of any new right (and rules on inheritance/succession) would need to be addressed. In many civil law jurisdictions, personality-based rights, such as the 'author's moral right of integrity' in France, are understood to be perpetual, subsisting even after the economic rights have expired. In the digital replica context, such a perpetual term seems excessive given the underlying objectives and concerns.

e) Data protection

Although the broader transparency obligation will lay the foundation for individuals to obtain more information about the processing of their biometric data (e.g., face and voice), it is overly optimistic to rely on individual rights to counter misuse. The weaknesses of informational self-determination and data protection have already been demonstrated.[74] Research has shown that around 32.5% of the 10,000 most popular UK's websites use implicit consent for the collection of personal data, and around 30.3%-56.2% of websites use pre-ticked options.[75] These frequently used methods of manipulation influence individuals' decision-making regarding the sharing and processing of their personal data. It would not be surprising to see similar practices being (or already having been) used to collect individuals' biometric data for digital replicas.

The unauthorised use of individuals' audiovisual data for digital replicas necessitates intervention at a systemic level, involving both regulatory authorities and platform operators. Focussing exclusively on individual rights may prove insufficient to address the risks on a broader scale and could result in unintended consequences, as evidenced by the challenges encountered with the GDPR. The high

72 Elena Cooper, 'AI and Performers' Rights in Historical Perspective' (2023) 45 EIPR 444; Laurence Bouvard, Elena Cooper and Amy Thomas, 'Evidencing the Value of Human Performance: Towards Re-Thinking Performers' Rights for an AI World' (2024) 46 EIPR 336.

73 *Irvine v TalkSport Ltd* [2002] EWHC 367 (Ch) and *Fenty v Arcadia Group Brands Ltd (trading as Topshop)* [2013] EWHC 1945 (Ch). See also Emma Perot, 'Anticipating AI: A Partial Solution to Image Rights Protection for Performers' (2024) 46 EIPR 407.

74 Wenlong Li and others, 'Mapping the Empirical Literature of the GDPR's (In-)Effectiveness: A Systematic Review' (2025) 57 Computer Law and Security Review 106129 <https://www.sciencedirect.com/science/article/pii/S2212473X25000021> accessed 28 May 2025.

75 Midas Nouwens and others, 'Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence' (2020) Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (ACM 2020) 1–13.

compliance costs and resulting barriers to data utilisation under the GDPR illustrate the potential for such frameworks to inadvertently hinder rather than facilitate the responsible use of data.

IV. Conclusion

This opinion has outlined one potential path forward for an innovation policy that balances rightsholder and creative industry positions while significantly enabling science and research. We suggest that changes to the current UK TDM exception (Sec. 29A CDPA 1988) could enable a wider and robust permissive space for early-stage R&D. By introducing transparency at market entry and an equitable remuneration requirement, an amended exception could also ensure that creatives can fairly benefit from new innovative uses of AI models trained on their works.

While tailored to the UK, the proposed intervention, in our view, contains the seeds for a global consensus. Our proposal is based on evidence relating to existing commercial agreements between rightsholders and AI developers for the training of models. Because of the need to access quality data and the legal uncertainty in major jurisdictions, there is a global move towards a licensing economy, regardless of the UK's policy choices. We also find that the earnings of creatives have declined dramatically over the last decade, and that income from the new kinds of licensing deals has not translated transparently into royalties for authors and performers.

Further evidence relates to the arrival of powerful open-source models that can be run locally rather than via platform services. On the intermediary side (platforms such as Ollama and Hugging Face) there is currently no clear liability regime for harms resulting from models made available.

Our proposal would achieve at least three goals. First it would remove obstacles for scientific research and education, recognising the specific imperatives of the humanities, social sciences and cultural sectors; second, transparency requirements at the point of market entry would clarify the licensing environment and facilitate access to quality content for R&D; finally, equitable remuneration rights would rebalance the bargaining framework and support creatives (authors, artists, performers).