HOUSE OF LORDS

Communications and Digital Committee

4th Report of Session 2024–26

AI, copyright and the creative industries

STRICTLY EMBARGOED UNTIL 00:01 FRIDAY 6 MARCH 2026. You must not disclose this document or its contents until the date and time above; any breach of the embargo could constitute a contempt of the House of Lords.

Ordered to be printed 24 February 2026 and published 6 March 2026

Published by the Authority of the House of Lords

HL Paper 267

Communications and Digital Committee

The Communications and Digital Committee is appointed by the House of Lords in each session "to consider the media, digital and the creative industries and highlight areas of concern to Parliament and the public".

Membership

The Members of the Communications and Digital Committee are:

Baroness Caine of Kentish Town
Viscount Colville of Culross
Baroness Elliott of Whitburn Bay
Baroness Fleet
Lord Holmes of Richmond
Baroness Keeley (Chair)
Lord Kirkhope of Harrogate
Lord Knight of Weymouth
Lord McNally
Baroness Owen of Alderley Edge
Lord Storey
Lord Tarassenko
The Lord Bishop of Winchester

Declaration of interests

See Appendix 1.

A full list of Members' interests can be found in the Register of Lords' Interests: https://members.parliament.uk/members/lords/interests/register-of-lords-interests

Publications

All publications of the Committee are available at: https://committees.parliament.uk/committee/170/communications-and-digital-committee/publications/

Parliament Live

Live coverage of debates and public sessions of the Committee's meetings are available at: http://www.parliamentlive.tv

Further information

Further information about the House of Lords and its Committees, including guidance to witnesses, details of current inquiries and forthcoming meetings is available at: http://www.parliament.uk/business/lords

Committee staff

The staff who worked on this inquiry were Eleanor Clements (Clerk), Ellie Halstead (Policy Analyst) and Rita Cohen (Committee Operations Officer).

Contact details

All correspondence should be addressed to the Communications and Digital Committee, Committee Office, House of Lords, London SW1A 0PW. Telephone 020 7219 2922. Email holcommunications@parliament.uk

Social media

You can follow the Committee on LinkedIn at https://www.linkedin.com/showcase/communications-and-digital-committee/.

Copyright and privilege

CONTENTS

Evidence is published online at https://committees.parliament.uk/work/9394/ai-and-copyright/

Q in footnotes refers to a question in oral evidence.

SUMMARY

The UK faces a choice between two futures. In the first, the UK becomes a world-leading home for responsible, licensing-based artificial intelligence (AI) development, where commercial model developers using UK content obtain permission, pay fair remuneration to rightsholders and can deploy their models without questions of legal liability. Domestic AI efforts would be directed towards building sovereign models whose training data and development processes are open to scrutiny. In this scenario, both the UK's creative industries and AI sector could thrive, building on our national strengths and unique selling point of innovating in creative technology.

In the second scenario, the UK continues to drift towards tacit acceptance of large-scale, unlicensed use of creative content and long-term dependence on opaque models trained overseas, with most benefits accruing to a small number of US-based firms while harms to UK creators grow.

Only the first path is compatible with the UK's long-term interests. The UK's creative industries are an economic powerhouse that contributed £124 billion to the UK economy in 2023, with gross value added expected to reach £141 billion by 2030. Their success is underpinned by a "gold-standard" copyright framework, which rewards creativity, supports sustainable business models for creative work, and commands international respect.

In the age of AI, the protections for creators afforded by copyright are under threat. Generative AI systems can now generate imitations of creative material in seconds, but speed is not a substitute for the value of the human creativity, skill and dedication that underpin original work. And these capabilities of AI systems depend on training models on vast quantities of human-created content, much of it copyrighted and drawn directly from the creative sector.

This is not because our copyright framework is outdated or in need of reform. Rather, widespread unlicensed use of protected works, coupled with limited transparency from AI developers about how their models have been trained, leaves rightsholders unsure about whether their content has been used, and unable to enforce their rights when it has. In addition, the absence of a robust 'personality right' or specific protection for digital likeness in the UK means creators and performers are unable to challenge harmful outputs that imitate their distinctive style, voice or persona.

These problems pose material risks to the livelihoods of individual rightsholders. Creators are already losing meaningful control over how their works and identities are used, leading to tangible economic harms, while an influx of AI-generated content in the market is replacing human-made work and undercutting paid commissions.

Meanwhile, technology sector stakeholders are pressing for the introduction in the UK of a broad new exception for commercial text and data mining (TDM) that would legitimise large-scale AI training on copyright-protected works. Without this, they argue, the growth of the UK's AI sector will be stunted.

There is, however, only limited evidence to show that weakening UK copyright law would significantly expand our AI sector. In contrast, a broad commercial TDM exception presents predictable harms to rightsholders by removing incentives to license protected works for AI training.

It would be a poor bet to sacrifice the UK's outstanding creative capacity for speculative AI gains. A new regime must now be created to safeguard creators' livelihoods, while harnessing the potential of AI for creativity and economic growth. To deliver this, we recommend the following actions:

- **Rule out a new commercial text and data mining exception with an opt-out model:** The Government was right to "reset" its initial approach to AI and copyright, but mixed public messaging and an extended consultation period have undermined trust and stalled licensing and investment. The Government should, in the next year, publish a final decision on its approach to AI and copyright. In the meantime, it should set out clearly that it will not introduce a new TDM exception with an opt-out mechanism.

- **Close gaps in protection for identity, style and digital replicas:** The Government should introduce protections against unauthorised digital replicas and harmful 'in the style of' AI outputs. These must give creators and performers clear control over commercial exploitation of their identity.

- **Make transparency about AI training data a statutory obligation:** The Government should establish a clear, mandatory transparency framework for AI developers. Any regime should be carefully designed to ensure disclosures are sufficiently granular to meet rightsholders' needs, while avoiding disproportionate burdens, particularly on small UK-based AI firms.

- **Create the conditions for a fair and inclusive UK licensing market:** A market for licensing content for AI use is already emerging. The Government should prioritise enabling a sustainable ecosystem that works for rightsholders and developers of different sizes, rather than relying on a single marketplace initiative such as its Creative Content Exchange pilot. As the market develops, the Government should also explore mechanisms to ensure that remuneration reaches individual creators, including through collaboration with existing collective management organisations (CMOs).

- **Champion the development of technical standards for control, provenance and labelling:** The Government must also back the creation and adoption of the technical tools that will support a licensing-first approach: open, interoperable and globally aligned standards for rights reservation, data provenance and the labelling of AI-generated content. It should be prepared to legislate where necessary to ensure effective implementation.

- **Prioritise the development and adoption of sovereign AI models:** International examples demonstrate that domestically governed AI systems can offer an alternative to an overreliance on opaquely trained US-based models. The Government's sovereign AI efforts should foster the creation of models that deliver enhanced transparency and respect for copyright.

AI, copyright and the creative industries

CHAPTER 1: INTRODUCTION

Background to our inquiry

1. Our aim in launching this inquiry was to feed into the Government's ongoing work on copyright and AI. We have sought to explore in more detail the issues raised by the Government in its consultation on possible reform of UK copyright law, which ran from December 2024 to February 2025. This was framed as addressing issues relating to copyright faced by the UK's creative industries and AI sector.[1] During the consultation period and around the debate on the Data (Use and Access) Bill that followed, views on reform became increasingly polarised. Discussions on this topic between the sectors continue to be fraught. We set out to examine the issues and, where possible, the technical tools that may help to address them, aiming to concentrate on the future, not the past.

2. In the light of the live policy debate on this topic, our inquiry programme was deliberately concise. We held seven oral evidence sessions in total, hearing from 21 witnesses. We also ran a targeted call for evidence and received 29 pieces of written evidence. We are extremely grateful to all those who contributed to our inquiry.

3. This Committee has examined AI and copyright several times since the previous Government announced a consultation on this in 2022.[2] Given our previous work, we recognise the complexity of this topic. We also understood that in the timeframe for this inquiry we would not uncover straightforward solutions. We hope, however, that this report will support constructive discussion on the issues surfaced by the Government's consultation, as well as informing future parliamentary debate on these and related areas.

The UK's creative industries and AI sector

4. In 2023, the UK's creative industries contributed £124 billion in GVA, and employed 2.4 million people.[3] They also deliver a strong export performance, with the UK holding a comparative advantage over countries such as the US, France and Germany.[4] The Government has identified the creative industries as having high-growth potential, with GVA expected to reach £141 billion in 2030.[5] Its Creative Industries Sector Plan identified that these sectors are "an economic success story … they play a unique role in both an economic

1 Intellectual Property Office, *Copyright and AI: Consultation*, 17 December 2024, p 2
2 See, for example, Communications and Digital Committee, *At risk: our creative future* (2nd Report, Session 2022–23, HL Paper 125), paras 26–47; *Large language models and generative AI* (1st Report, Session 2023–24, HL Paper 54), Chapter 8; *The future of news* (1st Report, Session 2024–26, HL Paper 39), paras 105–28; and *AI and creative technology scaleups: less talk, more action* (2nd Report, Session 2024–26, HL Paper 71), paras 172–79.
3 Written evidence from HM Government (AIC0013)
4 Creative Industries Policy and Evidence Centre, *UK Trade in a Global Creative Economy*, March 2024, pp 26–29
5 Written evidence from HM Government (AIC0013)

and social sense, and they are at the heart of our Industrial Strategy and our Growth Mission."[6]

5. In 2025, the Government set out two clear ambitions: enhancing the UK's position as a "global creative superpower";[7] and transforming it into "an AI maker, not an AI taker".[8] The Government's response to the *AI Opportunities Action Plan* emphasised the potential of AI to drive growth and productivity across the UK economy,[9] a view underlined in the *Digital and Technologies Sector Plan*.[10] According to the Government, the largest gains are expected to come from AI adoption across different sectors. In 2024, the GVA of "dedicated AI firms" reached £2.2 billion. The AI sector as a whole contributed £11.8 billion in GVA and employed around 86,000 people.[11]

Table 1: Size, GVA and employment figures for the creative industries and AI sector

	Creative industries (2023 figures)	AI sector (2024 figures)
GVA	£124 billion	£11.8 billion
Employment	2.4 million people	86,000 people

The Government's AI and copyright consultation

6. The Government's most recent consultation aimed to develop a "package of interventions" that would: support rightsholders to control the use of their content and seek remuneration for its use; enable "wide and lawful access to high-quality data" to support the development of "world-leading AI models in the UK"; and promote "greater trust and transparency" between the creative industries and the AI sector. It sought views on a series of possible ways forward:

   - maintaining the status quo (option 0);
   - strengthening copyright by requiring licences for the use of content in all cases (option 1);
   - creating a broad text and data mining (TDM) exception for commercial uses[12] (option 2); and

6 HM Government, *Creative Industries Sector Plan*, 23 June 2025, p 6
7 *Ibid.*, p 9
8 Department for Science, Innovation and Technology, *AI Opportunities Action Plan*, CP 1241, 13 January 2025
9 Department for Science, Innovation and Technology, *AI Opportunities Action Plan: government response*, CP 1242, 13 January 2025
10 HM Government, *Digital and Technologies Sector Plan*, 23 June 2025, pp 29, 35–37
11 Department for Science, Innovation and Technology, *Artificial Intelligence sector study 2024*, 3 September 2025
12 At present, UK copyright law includes an exception that permits TDM for non-commercial research purposes. See para 16.

- introducing a commercial TDM exception with a rights-reservation (opt-out) mechanism and associated transparency obligations,[13] mirroring elements of the European Union model[14] (option 3).

The Government also argued, however, that "the status quo cannot continue" and at that time presented option 3 as its "preferred" solution. Since then, there has been a "reset" and the Government now states that it no longer has a preferred option,[15] a point that was reiterated in evidence to our inquiry.[16]

Box 1: Text and data mining (TDM)

TDM refers to "the use of automated analytical techniques to analyse text and data for patterns, trends and other useful information." Typically, it "requires copying of the work to be analysed". UK copyright law currently includes an exception that permits copies to be made for TDM for "non-commercial research".

Source: Intellectual Property Office, Exceptions to copyright, 4 January 2021

The Data (Use and Access) Act 2025

7. In the same period, the Data (Use and Access) Bill was being considered in the House of Lords. Baroness Kidron put forward amendments to the Bill that sought to reinforce the application of existing copyright law in relation to AI models, and require greater transparency from developers about what copyrighted works were used in AI training.[17] The proposed changes gained support from members across the House and from high-profile figures from the creative sector.[18] A letter to the Prime Minister signed by over 400 creative industries stakeholders stated that the amendments would "put transparency at the heart of the copyright regime and allow both AI developers and creators to develop licensing regimes that will allow for human-created content well into the future."[19] However, AI founders, investors and some academics argued that the amendments posed "a serious threat to the UK's potential to lead in AI", and that the Government's consultation process was the appropriate route through which to develop "workable solutions".[20]

8. Introducing a debate about the amendments in the House of Commons, Sir Chris Bryant, then Minister for Creative Industries, Arts and Tourism, Department for Culture, Media and Sport, and Minister for Data Protection and Telecoms, Department for Science, Innovation and Technology, said:

13 The Government said that under this model, "AI developers would be able to train on material to which they have lawful access, but only to the extent that right holders had not expressly reserved their rights. It would mean that AI developers are able to train on large volumes of web-based material without risk of infringement. Importantly right holders are also able to control the use of their works using effective and accessible technologies and seek payment through licensing agreements". Intellectual Property Office, *Copyright and AI: Consultation*, p 13

14 See Table 2.

15 Culture, Media and Sport Committee, *Oral evidence: The work of the Department for Culture, Media and Sport*, 10 September 2025, Q 104 (Secretary of State for Culture, Media and Sport)

16 QQ 136, 138 (Secretary of State for Science, Innovation and Technology), 137, 150 (Secretary of State for Culture, Media and Sport)

17 HL Deb, 28 January 2025, cols 234–37

18 BBC News, *Peers demand more protection from AI for creatives*, 19 May 2025; The Times, *Elton John and Paul McCartney push for tougher AI copyright laws*, 10 May 2025 [paywall]

19 Music Publishers Association, *Creative industry leaders urge the Prime Minister to support UK creativity and economic growth by enforcing copyright law*, 12 May 2025

20 The Entrepreneurs Network, *Open letter: Data (Use and Access) Bill*, 12 May 2025

> "We want genuine transparency about what is used in training AI, alongside rightsholders' control of their work and appropriate access to training material for AI. However, although I accept the intention behind the amendment … we do not believe that the Bill, which is a data measure, is the right vehicle for action."[21]

9. Baroness Kidron's amendments were ultimately not adopted. There were, however, some compromises. The Data (Use and Access) Act (DUAA) states that by 19 March 2026, the Secretary of State must:

    - Publish an assessment of the economic impact of each of the four policy options described in the Government's consultation paper, including the impact on copyright owners, developers and users.[22]
    - Publish a report on the use of copyright works in the development of AI systems. This must address :
        - technical measures and standards that may be used to control the use of works to develop AI systems;
        - the effect of copyright on access to and use of data by developers;
        - the disclosure of information by developers about their use of copyright works to develop AI systems;
        - the granting of licences to AI developers to do acts restricted by copyright; and
        - ways of enforcing requirements and restrictions relating to the access and use of copyright works.[23]

10. The DUAA also required the Government to publish an update on its progress in December 2025.[24] This set out that that the consultation had received more than 11,500 responses, including 10,112 submitted via the Citizen Space online survey. Among Citizen Space respondents, the overwhelming majority of responses (88%) supported licensing in all cases (option 1). Only 3% supported the introduction of a new commercial TDM exception accompanied by a rights-reservation mechanism and transparency measures (option 3). Just 0.5% of respondents supported a broader commercial TDM exception with no rights-reservation at all (option 2).[25]

Focus of our report

11. In the light of the Government's forthcoming publications under the DUAA, and the topics covered by the technical working groups it established in December,[26] our report focuses on the following issues:

    - Chapter 2 examines the UK's existing copyright framework, stakeholder perspectives on its application to generative AI systems and the Government's handling of work in this area to date;

21 HC Deb, 11 March 2025, cols 84–85
22 Data (Use and Access) Act, section 135
23 Data (Use and Access) Act, section 136
24 Data (Use and Access) Act 2025, section 137
25 See paras 166–67 for further discussion of views expressed in the consultation responses. HM Government, *Copyright and artificial intelligence statement of progress under Section 137 Data (Use and Access) Act*, 15 December 2025, pp 10–11
26 *Ibid.*, pp 12–13

- Chapter 3 examines why transparency with regard to training data has emerged as a central issue in debates on AI and copyright, key considerations for potential future transparency requirements and the opportunities presented by sovereign AI models;
- Chapter 4 examines technical controls that could allow rightsholders to express preferences about how their content is used by AI systems, as well as labelling and provenance technologies; and
- Chapter 5 examines the emerging market for licensing content for generative AI use, and how the opportunities this presents could be maximised.

CHAPTER 2: AI, THE CREATIVE INDUSTRIES AND COPYRIGHT REFORM

12. In this chapter, we set out the existing landscape in relation to AI and copyright in the UK, before assessing the case for change.

Copyright and the generative AI lifecycle

The UK's existing copyright framework

13. Generative AI refers to systems that can produce content across multiple modalities, such as text, images, audio, video and computer code, in response to user prompts.[27] These systems are developed using text and data mining (TDM) techniques that involve the copying and automated analysis of vast quantities of data to identify statistical patterns and relationships.[28] The resulting model can then use those patterns to generate outputs that may closely resemble human-created content.[29]

14. Much of the material used in developing and operating generative AI systems consists of human-created works protected by the Copyright, Designs and Patents Act 1988 (CDPA). These include books, news articles, computer programs, photographs, musical works, films and broadcasts.[30] Other material, such as raw factual data or public domain content,[31] may fall outside copyright protection.

15. Copyright arises automatically in qualifying works[32] and gives rightsholders exclusive control over a number of 'restricted' acts, including copying.[33] Copyright underpins the UK's creative economy by granting creators exclusive rights to use and distribute their works for a defined period, incentivising the creation and circulation of new works by enabling creators to control how their material is used and to seek payment for their creative

27 Parliamentary Office of Science and Technology, *Artificial intelligence and new technology in creative industries*, 7 October 2024. Examples of generative AI models include text-generation systems such as large language models (for example, OpenAI's ChatGPT and Google's Gemini), image and multimodal models for visuals (for example, Stable Diffusion), models specialised for video synthesis (for example, Open AI's Sora and Google's Veo 3), audio and music generation (for example, ElevenLabs), and code generation models (such as GitHub Copilot).

28 House of Commons Culture, Media and Sport Committee, *Connected tech: AI and creative technology* (11th Report, Session 2022–23, HC Paper 1643), pp 12–13

29 Intellectual Property Office, *Copyright and AI: Consultation*, p 23

30 Copyright, Designs and Patents Act 1988, section 1

31 Material can enter the public domain if the author dedicates it to the public domain or the copyright has expired, allowing anyone to use it freely without needing permission. Coventry University Library, *The Public Domain*, 17 September 2025

32 To qualify for copyright under the UK Copyright, Designs and Patents Act 1988 (CDPA), a work must fall within one of the statutory categories and satisfy the originality threshold. The primary categories of qualifying works include original literary (expressly encompassing computer programs and software code), dramatic, musical and artistic works; sound recordings; films; broadcasts; and typographical arrangements of published editions. Copyright, Designs and Patents Act 1988, section 1(1). Originality under UK law is assessed by reference to the author's own intellectual creation, requiring that the work originates from human creative choices rather than mere copying from existing sources. Traditionally, this test is satisfied where the author has applied skill, judgment and individual effort. Unlike most jurisdictions, the CDPA contains a separate provision for "computer-generated works", defined as works generated by a computer in circumstances such that there is no human author. Copyright, Designs and Patents Act 1988, section 9(3). Ideas are not protected by copyright; only the expression of those ideas as fixed in material form are protected. Pinsent Masons, *UK Copyright law: the basics*, 21 June 2024

33 Restricted acts include the right to copy, issue copies, rent or lend, perform, show, play or communicate the work to the public, or adapt it. Copyright, Designs and Patents Act 1988, section 16

effort.[34] Several witnesses to our inquiry characterised the UK's copyright framework as the international "gold standard".[35]

16. Copyright may be infringed if the whole or a substantial part of a protected work is copied without permission, unless an exception applies.[36] The CDPA provides two exceptions that are worth highlighting in this context:

- Section 29A permits copies to be made for TDM for the sole purpose of research for non-commercial purposes, provided the content is lawfully accessed and, where practical, sufficiently acknowledged. This is the only exception that is expressly directed at TDM.[37]
- Section 28A is a more general exception. This allows temporary copies of works to be made as part of a technological process, where those copies are transient or incidental, form an integral and essential part of the process, have no independent economic significance, and exist solely to enable a lawful use of the work.[38]

There is currently no broad exception that explicitly permits the use of copyrighted works for commercial TDM or the training of generative AI models, unlike in some other jurisdictions (see Table 2).[39]

Table 2: Copyright provisions for text and data mining (TDM) in selected jurisdictions

Jurisdiction	Provision	Summary
United Kingdom	Copyright, Designs and Patents Act 1988, Section 29A	A specific statutory exception permitting TDM for non-commercial research only, provided the user has lawful access to the work.[40]
European Union	Copyright in the Digital Single Market Directive (EU) 2019/790, Articles 3 and 4	Article 3 permits TDM by research organisations and cultural heritage institutions for scientific research. Article 4 permits TDM of lawfully accessible works for both commercial and non-commercial purposes, unless rightsholders have expressly reserved their rights (e.g. via machine-readable means).[41]

34 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008). See also Intellectual Property Office, *Copyright and AI: Consultation*, p 9.
35 QQ 2 (Isabelle Doran, Anna Ganley, Owen Meredith), 42, 50 (Serena Dederding), written evidence from Independent Society of Musicians (AIC0014), written evidence from Copyright Licensing Agency (AIC0005) and supplementary written evidence from News Media Association (AIC0022)
36 Intellectual Property Office, *Enforcing your copyright*, 29 May 2014
37 Copyright, Designs and Patents Act 1988, section 29A
38 Copyright, Designs and Patents Act 1988, section 28A
39 See Table 2.
40 Copyright, Designs and Patents Act 1988, section 29A
41 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, OJ L 130/92, 17 May 2019

Jurisdiction	Provision	Summary
Japan	Copyright Act, Article 30–4	A broad statutory exception allowing TDM for commercial and non-commercial uses, particularly for "non-enjoyment purposes".[42]
Singapore	Copyright Act 2021, Section 244	A statutory exception permitting the copying of copyrighted works for computational data analysis purposes, including TDM, applicable to both commercial and non-commercial uses.[43]
United States	Copyright Act 1976, §107 ("fair use")	No specific TDM provision. Use of copyrighted works, including for AI training, is assessed under the fair use doctrine using a four-factor test applied case by case.[44]

AI model development and operation

17. Because TDM and other stages of developing and operating generative AI systems involve copying and processing protected works,[45] copyright may be engaged at several points in the 'lifecycle'[46] of a model. Witnesses to our inquiry commonly distinguished between acts of copying and related processing that occur when material is collected, prepared, analysed or retrieved for use by an AI system (often referred to as 'input' concerns); and any later reproduction of protected expression in AI-generated outputs.[47] Dr Hayleigh Bosher, Associate Professor in Intellectual Property Law at Brunel, University of London, and Dr Andres Guadamuz, Reader in Intellectual Property Law at the University of Sussex, identified four principal stages of the generative AI lifecycle at which copyright may be engaged: data collection and pre-processing; model training; inference; and output generation.[48] These stages are set out in Figure 1.

42 Matthew Sag and Peter K Yu, 'The Globalization of Copyright Exceptions for AI Training', *Emory Law Journal*, vol 74, Issue 5, (2025), pp 1185–86: https://scholarlycommons.law.emory.edu/cgi/viewcontent.cgi?article=1562&context=elj

43 *Ibid.*, pp 1190–92

44 To determine whether the fair use exception applies in a copyright infringement lawsuit, four factors are considered: (1) the purpose and character of the use, including whether it is "transformative"; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect on the potential market for or value of the copyrighted work. See U.S. Copyright Office, *U.S. Copyright Office Fair Use Index*, August 2025

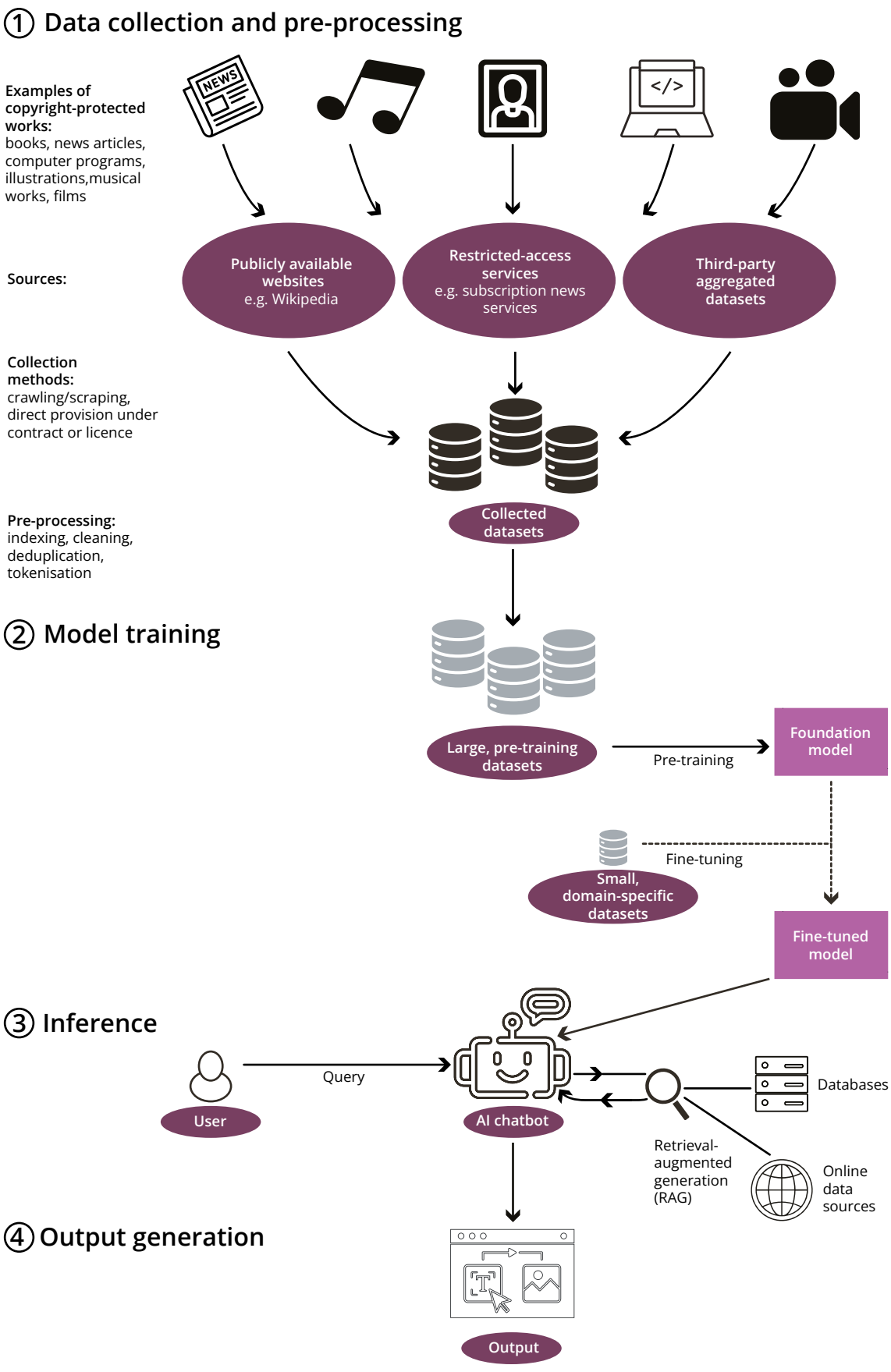
45 Written evidence from Dr Hayleigh Bosher (AIC0025) and Dr Andres Guadamuz (AIC0026)

46 We use the term 'lifecycle' to refer to "the end-to-end process of developing, deploying and managing AI systems. It is an ongoing cycle that includes stages such as data collection, data processing, model training, evaluation, deployment, monitoring and governance". João Pedro Quintais, 'Copyright, the AI Act and Extraterritoriality', *Policy Brief—The Lisbon Council* (June 2025), p 4: https://dx.doi.org/10.2139/ssrn.5316132. See also Q 80 (Prof John Collomosse) and written evidence from Dr Andres Guadamuz (AIC0026).

47 QQ 63 (Matthew Sinclair), 80 (Eugene Huang), 88 (Ed Conolly), written evidence from Meta (AIC0009), written evidence from Google (AIC0010), written evidence from techUK (AIC0011), written evidence from HM Government (AIC0013) and written evidence from Directors UK and British Equity Collecting Society (AIC0018)

48 Written evidence from Dr Hayleigh Bosher (AIC0025) and Dr Andres Guadamuz (AIC0026)

Figure 1: The generative AI lifecycle



Model development

- **Stage 1—Data collection and pre-processing:** At this stage, developers assemble large collections of copyright-protected and other material from publicly available websites, restricted-access services and other public datasets.[49] Content may be obtained via automated 'bots' that discover and retrieve web pages (a process known as 'crawling'), and extract and copy specific content from those pages into separate databases (known as 'scraping').[50] Alternatively, they may be supplied directly under contract or licence. Multiple copies of protected works are typically created, even if only temporarily, to enable indexing, cleaning, deduplication and conversion into machine-readable or tokenised formats[51] that may no longer be recognisable as the original expression.[52] The main legal questions here are whether these acts of copying and transformation are authorised by licence or exception,[53] and whether the material has been sourced lawfully (for example, licensed content as opposed to pirated or paywall-circumventing copies).[54]

- **Stage 2—Model training:** In the training stage, the prepared datasets are iteratively fed as inputs to machine learning models whose internal parameters, known as 'weights', are repeatedly adjusted so that the model captures statistical patterns and relationships in the input data. The result is a set of numerical parameters rather than accessible copies of individual works.[55] However, models may in some cases 'memorise' and later reproduce particular passages, images or other content.[56] Different datasets may be used for initial pre-training and for fine-tuning the model for specific tasks or domains. Key legal issues include whether the copying and intermediate processing of the data

---

49 For example, a 2020 paper described how OpenAI trained GPT-3 on a mixture of filtered Common Crawl web data, two internet-based books corpora (Books1 and Books2), and Wikipedia. Tom B Brown et al, 'Language Models are Few-Shot Learners', *Proceedings of the 34th International Conference on Neural Information Processing Systems*, (2020): https://doi.org/10.48550/arXiv.2005.14165. Other academics noted that training data is drawn from freely available online materials (for example, Project Gutenberg and Spoken Wikipedia), licensed corpora and subscription databases (for example, scholarly literature held in databases to which developers have access). Martin Kretschmer, Thomas Margoni and Pinar Oruç, 'Copyright Law and the Lifecycle of Machine Learning Models', *IIC*, vol 55 (2024), 110–38: https://doi.org/10.1007/s40319–023-01419-3. In a 2025 paper, the OECD distinguished between targeted datasets (including open resources and protected corpora obtained under licence) and large-scale scraped data sourced from books, websites, blogs, forums, databases and social media platforms. OECD, 'Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data', *OECD Artificial Intelligence Papers*, no 33 (2025): https://doi.org/10.1787/d5241a23-en

50 Cloudflare, *What is a web crawler? How web spiders work*

51 Written evidence from Dr Hayleigh Bosher (AIC0025) and Dr Andres Guadamuz (AIC0026)

52 "It is this extensive cleaning process that makes the copyright argument for individual works more difficult as well, works are no longer recognisable in their original format, they're data that is put together with other similar data." TechnoLlama (Dr Andres Guadamuz), *Revisiting copyright infringement in AI inputs and outputs*, 30 July 2025

53 See para 16.

54 In the US case of *Bartz v Anthropic*, the AI company Anthropic initially trained successive large language models on unauthorised copies of books, from so-called shadow libraries, to build a central library. This was not considered 'fair use' as the source material was illegally obtained. As a result, Anthropic faced substantial damages for copyright infringement, which led to a $1.5 billion settlement for authors of the unauthorised books. Inside Tech Law, *Bartz v. Anthropic: Settlement reached after landmark summary judgment and class certification*, 5 September 2025

55 Arne Radeisen, 'Open Foundation Models and TDM Exceptions to Copyright—Building Blocks for an AI Ecosystem', *GRUR International*, vol 75, Issue 3, (2026), p 233: https://doi.org/10.1093/grurint/ikag002

56 TechnoLlama (Dr Andres Guadamuz), *Revisiting copyright infringement in AI inputs and outputs*, 30 July 2025

required for training, even if it is transient, tokenised, or later deleted,[57] are authorised by licences or exceptions, and the extent to which any tendency of the model to 'memorise'[58] specific works increases the risk that later outputs will reproduce a substantial part of a protected work.

Model operation

- **Stage 3—Inference:** Once trained, models are integrated into products and services such as chatbots, search and summarisation tools, and creative applications. User prompts are converted into numerical form and passed through the model to generate a response. Many services supplement this process with retrieval-augmented generation (RAG), where the system retrieves and incorporates external, up-to-date information (such as recent news stories) at the point of use.[59] This can involve further copying, caching and processing of protected works beyond the training stage.[60] The system determines the content of its response from a combination of the user's prompt, any retrieved reference materials and the patterns encoded in the model's parameters. Because the internal workings of models are difficult to interpret, it is challenging in practice to determine how particular outputs draw on specific training or retrieved information.[61]

- **Stage 4—Output generation:** The result is an output presented to the user in the form of text, images, audio or other media. In most cases, these outputs may be considered novel compositions that synthesise patterns derived from many different sources.[62] From a copyright perspective, the central concern is where an output reproduces a substantial part of a particular protected work and communicates it to the public, thereby engaging two copyright-restricted acts[63]—for example, where 'memorised' text or images are recreated at a user's request, or where retrieved material is quoted or displayed. Outputs that merely mimic a general style without reproducing specific protected expression will not usually infringe copyright, although they may still substitute for human-created works in practice and negatively impact rightsholders in other ways, as we explore later in this chapter.[64]

---

57 Daniel J Gervais, Noam Shemtov, Haralambos Marmanis, Catherine Zaller Rowland, 'The Heart of the Matter: Copyright, AI Training, and LLMs', (2024), pp 9–13: https://dx.doi.org/10.2139/ssrn.4963711

58 "Memorisation in machine learning refers to the capacity of a model to retain and reproduce specific items from its training data rather than merely learning generalised patterns". Written evidence from Dr Andres Guadamuz (AIC0026). See also A. Feder Cooper et al, 'Extracting memorized pieces of (copyrighted) books from open-weight language models', (2025): https://doi.org/10.48550/arXiv.2505.12546; Ahmed Ahmed, A. Feder Cooper, Sanmi Koyejo, Percy Liang, 'Extracting books from production language models', (2026): https://doi.org/10.48550/arXiv.2601.02671.

59 Shailja Gupta, Rajesh Ranjan, Surya Narayan Singh, 'A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions' (2024): https://doi.org/10.48550/arXiv.2410.12837

60 European Parliament, *Generative AI and Copyright—Training, Creation, Regulation*, July 2025, pp 45– 50

61 Communications and Digital Committee, *Large language models and generative AI* (1st Report, Session 2023–24, HL Paper 54), para 13

62 However, there may be instances where a model appears to regurgitate its training data. See Stage 2 above.

63 Intellectual Property Office, *Copyright and AI: Consultation*, pp 10–11

64 See paras 40–45.

Arguments for and against copyright reform

Use of protected works in AI training

18. Much of the current policy debate concerns how copyright applies to the use of protected works in training generative AI models, particularly at the data collection and model training stages, rather than on questions of infringement in AI-generated outputs. Originally, in its December 2024 consultation on AI and copyright, the Government characterised the present position as one of "legal uncertainty" for AI developers. It suggested that some leading developers therefore choose to train models in jurisdictions with clearer or more permissive rules, a point that was echoed in our evidence.[65] The consultation sought views on whether a broad new commercial TDM exception should be introduced and initially said that an 'opt-out' approach was its preferred route.[66] Since then, there has been a "reset" and the Government now states that it has no preferred option.[67]

19. Determining whether training generative AI models on copyrighted works without a licence infringes the reproduction right is ultimately a matter for the courts. There has been no ruling in the UK on this specific question to date,[68] and courts in other jurisdictions are still working through these questions under their own doctrines.[69] In the absence of UK case law, the Government's consultation suggested that it may be appropriate to consider "a more direct intervention through legislation to clarify the rules in this area".[70]

20. During our inquiry, creative industry witnesses fundamentally disagreed with the premise that the application of the current copyright framework to AI development is uncertain. They were unanimous that UK copyright law is "absolutely clear"[71] on this point: in their view, copying protected works to train AI models engages the restricted act of reproduction and therefore requires rightsholders' consent and a licence under the CDPA, unless a

65 Q 61 (Vinous Ali), written evidence from OpenAI (AIC0004), written evidence from CREATe Centre, University of Glasgow (AIC0007), written evidence from Meta (AIC0009), written evidence from Google (AIC0010) and written evidence from Dr Andres Guadamuz (AIC0026)

66 Intellectual Property Office, *Copyright and AI: Consultation*, p 10

67 QQ 136 (Secretary of State for Science, Innovation and Technology), 150 (Secretary of State for Culture, Media and Sport)

68 The High Court judgment in the case between the image library Getty Images and the AI company Stability AI in November 2025 did not decide whether the act of training a model on copyright-protected works without a licence infringes the reproduction right. Shortly before closing submissions, Getty Images abandoned its primary copyright claim, having accepted that there was no evidence that Stability AI's Stable Diffusion model had been trained or developed in the UK. The Court therefore addressed only the secondary infringement claim—that is, whether the model weights made available in the UK were themselves an "infringing copy". This aspect of the claim failed. Browne Jacobson, *Getty Images' copyright not infringed by Stability AI making its Stable Diffusion model available to users in the UK*, 12 November 2025. In December 2025, Getty Images was given permission to appeal against the decision on secondary copyright infringement on the basis that this raised a "novel and important" point of law. CMS Law-Now, *Getty Images v Stability AI - permission to appeal sought by Getty Images at consequentials hearing*, 21 January 2026

69 See Table 2.

70 Intellectual Property Office, *Copyright and AI: Consultation*, p 10

71 Written evidence from Creative Rights in AI Coalition (AIC0001)

specific statutory exception applies.[72] The Independent Society of Musicians (ISM) told us:

> "In principle, UK copyright law already gives rightsholders the exclusive rights they need to control the use of their works, including reproduction, communication to the public, and adaptation. The problem is not the law's scope, but the lack of effective enforcement in the AI context."[73]

21. Dr Bosher emphasised that the restricted act of reproduction is "a broad legal concept", giving rightsholders the exclusive right to control "direct or indirect, temporary or permanent reproduction by any means, in any form, in whole or in part". In her view, "it does not matter so much *how* a copy is made, it is simply enough that a protected work is used without permission", because copyright "regulates copying, in any material form".[74] Dr Guadamuz similarly told us that "what copyright does protect, and what AI training necessarily engages, is the reproduction" of copyrighted material, since "the act of gathering, copying, and processing works to build a training dataset will, in almost every case, involve the making of a copy of a work at some point in the process". On that basis, he concluded that "the unauthorised copying of protected works for AI training is capable of constituting infringement" under the CDPA.[75]

22. Turning to whether the existing copyright exceptions permit the copying involved in AI training, the Centre for Regulation of the Creative Economy at the University of Glasgow (CREATe) commented that in its view the "narrowly conceived" non-commercial research exception makes research and development (R&D) in the UK "risky". It proposed that section 29A of the CDPA should be reconceived to cover "all research and development before market entry, at which point transparency and licensing obligations apply".[76] On a similar note, Google and Microsoft told us that the exception is not suited to contemporary AI development practices. It is, they argued, increasingly difficult to distinguish non-commercial TDM from commercially linked activity, given that advanced AI development is frequently achieved through collaboration between universities, public sector bodies and the private sector, and where the end goal is often commercialisation.[77]

23. The News Media Association (NMA) disputed strongly that the current non-commercial research exception leaves a problematic gap for AI research and development. In its view, the constraints on onward sharing of datasets are a deliberate safeguard to prevent research corpora becoming a backdoor into unlicensed commercial model training; the perceived risk to public-private partnerships is therefore a "feature, not a bug" of the law. The NMA stated that extending section 29A to all pre-market R&D would primarily benefit large commercial AI developers by allowing models to be trained on copyright works without consent "at the very stage when rightsholders' leverage matters most", undermining the emerging licensing market by

72 Written evidence from Alliance for Intellectual Property (AIC0002), written evidence from Copyright Licensing Agency (AIC0005), written evidence from British Copyright Council (AIC0006), written evidence from Independent Society of Musicians (AIC0014), written evidence from Music Publishers Association (AIC0017), written evidence from Dr Hayleigh Bosher (AIC0025) and supplementary written evidence from Society of Authors (AIC0003) and News Media Association (AIC0022)
73 Written evidence from Independent Society of Musicians (AIC0014)
74 Written evidence from Dr Hayleigh Bosher (AIC0025)
75 Written evidence from Dr Andres Guadamuz (AIC0026)
76 Written evidence from CREATe Centre, University of Glasgow (AIC0007)
77 Written evidence from Google (AIC0010) and Microsoft (AIC0012)

removing incentives to seek permission up front. It also noted that relying on an undefined concept of "market entry" would make it difficult for rightsholders to monitor or enforce their rights.[78]

24. Dr Bosher stated in simple terms that the current exception for TDM in section 29A "does not apply to AI models because it is only applicable for non-commercial purposes". She also rejected reliance on the temporary copying exception in section 28A, noting that she does not consider its associated conditions to be met by large-scale generative model training. If any existing exception already authorised these uses, she argued, the Government "would not have considered implementing a new copyright exception to cover AI activity".[79]

25. Input to our inquiry from technology sector witnesses generally focused on whether using copyright-protected content for model training involves the kind of 'copying' that the CDPA is intended to regulate, rather than the detail of existing exceptions. Some witnesses argued that training models on publicly available content should be understood as a process analogous to 'learning' patterns from information, rather than copying expressive content in a way that engages the reproduction right.[80] Dr Guadamuz recognised the "intuitive appeal" of the analogy: "We do not say that a novelist infringes copyright merely because they have read and been influenced by other novels. If this process of learning from existing works does not constitute infringement when performed by a human, should it when performed by a machine?"[81] Matthew Sinclair, Senior Director at the Computer and Communications Industry Association (CCIA), proposed that, as such, licensing should only be considered "when copyright is used in the outputs of models" or where "distinctive datasets" are concerned.[82]

26. Witnesses also drew attention to the scale and composition of training data. Google argued that "any particular work in and of itself is not necessary" to train a model; these rely instead on the "total collection of works".[83] Meta similarly characterised training as extracting "non-expressive data and grammatical patterns … from the underlying training data",[84] with no individual work having a "material influence" on the model as a whole.[85] Mr Sinclair stated that as a result, the contribution of any given piece of content to a large training corpus is "in effect, zero",[86] and that material with no identifiable economic value at the level of individual works should not, in itself, be subject to a separate licensing requirement.

27. A further contention was that generative AI systems are designed to produce novel outputs, not to reproduce the works on which they were trained. Roxanne Carter, Global IP Lead for the Government Affairs and Policy team at Google, told us that since systems are intended to "produce wholly new content", rather than to function as an "information retrieval system" or

78 Further supplementary written evidence from News Media Association (AIC0029)
79 Written evidence from Dr Hayleigh Bosher (AIC0025)
80 QQ 56–57, 59 (Matthew Sinclair), 57 (Vinous Ali)
81 Written evidence from Dr Andres Guadamuz (AIC0026)
82 Q 56 (Matthew Sinclair)
83 Written evidence from Google (AIC0010)
84 In copyright law, 'expression' refers to the original form in which ideas are presented, that is "the artistic expression of an idea or fact, not the fact or idea itself". Intellectual Property Office, *Exceptions to copyright: Research*, October 2014
85 Written evidence from Meta (AIC0009)
86 Q 58 (Matthew Sinclair)

databases of underlying works, developers should be free to train on material that is freely available on the web without obtaining licences at that stage.[87] However, we note that this distinction does not reflect how Google's products operate in practice. Several of its AI tools, including Google Gemini, Google AI Overviews and Google AI Mode, integrate RAG to surface, summarise and organise information from specific sources at the point of use, and routinely direct users to those sources through hyperlinks. In these services, retrieval and generative functions are deliberately combined.[88]

28. Recent official assessments and academic commentary in other jurisdictions have, however, cast doubt on the 'non-expressive use' and human-learning analogies advanced by some developers. The US Copyright Office, for example, has rejected the claim that generative AI training is non-expressive once datasets of protected works are used to build models capable of reproducing copyrighted expression. It concluded that the learning analogy "rests on a faulty premise": humans retain only imperfect, filtered impressions of works, whereas generative AI training involves making perfect digital copies, analysing them almost instantaneously, and producing models that can create competing works at "superhuman speed and scale".[89]

29. Dr Guadamuz similarly emphasised "the industrial scale of the process" of AI training, which "distinguishes it from anything that occurs in human learning". He concluded that we should analyse AI training "through the established doctrines of copyright law … rather than seeking to import a metaphor that was never designed to bear legal weight".[90] Dr Bosher characterised terms such as 'memorisation', 'reading' and 'learning' as "misleading metaphorical rhetoric" that is "deliberately used to influence the way in which the functioning of AI is understood". She argued that while these "might be accepted parlance in the AI sector", they should be "avoided in the context of policy and regulation" as they obscure the underlying technical processes and their legal and economic significance.[91]

Calls for a commercial TDM exception

30. Another common theme in evidence from technology sector representatives was that "legal uncertainty"[92] arises from the UK's current copyright framework, which then acts as a barrier to AI investment. Vinous Ali, Deputy Executive Director at the Startup Coalition, described the framework as a "ceiling to our ambition", telling us:

> "We can have the best research but that will not be enough. We can have the Government pouring £2 billion into AI and unblocking planning laws, infrastructure and energy costs, but that will not be enough if start-ups do not have access to the data they need to create."[93]

87 Q 107 (Roxanne Carter)
88 Google explains RAG as an AI framework that "combines the strengths of traditional information retrieval systems (such as search and databases) with the capabilities of large language models", and discusses using Gemini with search-based grounding. Google Cloud, *What is Retrieval-Augmented Generation (RAG)?*. Professor Chirag Shah of the University of Washington explains that Google's AI Overviews is a RAG system that retrieves top results and then uses them to generate answers. Tech Policy Press, *What Does It Take To Moderate AI Overviews?*, 22 September 2025
89 U.S. Copyright Office, *Copyright and Artificial Intelligence, Part 3: Generative AI Training Pre-Publication Version*, May 2025, p 46
90 Written evidence from Dr Andres Guadamuz (AIC0026)
91 Written evidence from Dr Hayleigh Bosher (AIC0025)
92 Written evidence from Meta (AIC0009) and Microsoft (AIC0012)
93 Q 65 (Vinous Ali)

Mr Sinclair agreed that the Government's efforts to "address issues such as energy costs and skills" could be "undermined by not having the right copyright regime."[94]

31. These stakeholders consistently identified a broad commercial TDM exception, with or without the opt-out mechanism proposed by the Government in its consultation, as the central policy lever to address these concerns.[95] Google told us that without a TDM exception for commercial use, "the UK will struggle to achieve its stated ambitions in becoming an attractive place for AI investment and innovation, as training AI models in the UK will be more difficult for companies of all sizes."[96]

32. Meta cited analysis by Europe Economics and the CCIA[97] which estimated that if the UK had a commercial TDM exception, it would secure $6.4 billion of AI investment overall in 2025, of which $4.3 billion would be "TDM-sensitive". Without a TDM exception, the overall AI investment figures were estimated to be between $1.9 billion and $3.2 billion.[98] As part of this analysis, a survey of developers, investors, and others in the UK AI sector was conducted. It found that if other jurisdictions implemented protections for TDM and the UK did not, 66% of respondents would undertake projects in other countries, and 25% reported that they would have to do this for "many" projects.[99] However, we note that the Europe Economics analysis explicitly states that it is focused on only "one specific metric: investment in TDM-intensive activities".[100] It therefore does not address the corresponding impact that introducing a commercial TDM exception might have on the creative industries, which are already experiencing significant AI-related harms, as we explore later in this chapter.[101]

33. When asked to rank the significance of copyright as a barrier, in comparison to issues such as energy prices and access to capital, Antony Walker, Deputy Chief Executive Officer of techUK, told us that it is "of pretty equal relevance".[102] In contrast, the Creative Industries Policy and Evidence Centre (Creative PEC) suggested that there is currently limited robust evidence on the precise impact of copyright legislation on AI investment decisions. In its view, there is "little evidence that copyright is the main reason" dedicated AI companies or startups choose to base key operations elsewhere.[103] Ed Newton-Rex, Chief Executive Officer of Fairly Trained, noted that the trend for the largest technology companies to base themselves outside the UK predates generative AI. The reasons why, he argued, "have nothing to do with copyright law".[104]

---

94 Q 64 (Matthew Sinclair)
95 QQ 58 (Antony Walker), 58, 68 (Vinous Ali), written evidence from OpenAI (AIC0004), written evidence from Google (AIC0010), written evidence from Meta (AIC0009) and written evidence from Microsoft (AIC0012)
96 Written evidence from Google (AIC0010)
97 Written evidence from Meta (AIC0009)
98 Europe Economics, *Estimating the impact on investment of a commercial TDM exception*, June 2025, p 1
99 *Ibid.*, pp 6–7
100 *Ibid.*, p 5
101 See paras 40–45.
102 Q 65 (Antony Walker)
103 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)
104 Q 51 (Ed Newton-Rex)

34. In this context, we note that the Government's *AI Opportunities Action Plan* made only limited reference to copyright.[105] It set out a broader agenda for strengthening the industry, including developing a long-term plan for AI infrastructure, establishing AI Growth Zones to accelerate data-centre development, and improving access to compute.[106] The Government recently reported that 38 of its 50 commitments had been met in the first year of the plan's implementation.[107] techUK said there had been "notable progress in foundational areas", including compute planning, skills programmes and the establishment of the Sovereign AI Unit.[108] The update also highlighted progress on unlocking high-value creative and research data through initiatives such as the Creative Content Exchange,[109] but did not frame copyright as a material barrier to delivery.

35. In our inquiry on scaling up, access to data was raised as a factor that can limit firms' growth. Commentary focused, however, on facilitating access to key public data assets and targeted datasets, rather than the broad range of content that a new TDM exception would cover. Challenges with accessing finance, compute and talent were viewed as more significant blockers.[110] Similarly, in the Government's Artificial Intelligence Sector Study 2024, concerns about regulatory clarity were mentioned less frequently than availability of finance and talent as barriers to firm growth. The study also highlighted the growth which has occurred in the UK AI sector with the current copyright framework and without a commercial TDM exception.[111]

36. **The interpretation of UK copyright law is a matter for the courts and future cases may help to set precedents in this area. However, the consistent call from technology sector stakeholders for a new, broad commercial text and data mining (TDM) exception expressly to enable more AI model training to take place in the UK suggests that they do not regard large-scale commercial training on copyright-protected works as clearly covered by the existing exceptions. If they did, a commercial TDM exception would be unnecessary.**

37. **On this basis, the main uncertainty for large AI developers appears to lie in the question of whether their current and proposed training practices would withstand legal challenge if tested in court. Support for a broad commercial TDM exception should therefore be understood as an attempt to lower that litigation risk by weakening the current level of copyright protection, rather than as a neutral exercise in clarifying the law. We are also not persuaded that expanding the existing non-commercial research exception to cover all "pre-market" research and development is either necessary or desirable.**

---

105 Department for Science, Innovation and Technology, *AI Opportunities Action Plan*, 13 January 2025, CP 1241

106 Compute is defined as "the systems assembled at scale to tackle computational tasks beyond the capabilities of everyday computers". It includes physical supercomputers and cloud computing. Parliamentary Office of Science and Technology, *Artificial intelligence (AI) glossary*, 23 January 2024

107 Department for Science, Innovation and Technology, *AI Opportunities Action Plan: One Year On*, 29 January 2026

108 techUK, *Delivery must now be the focus of the UK's AI Opportunities Action Plan in 2026*, 30 January 2026

109 See paras 231–39.

110 Communications and Digital Committee, *AI and creative technology scaleups: less talk, more action* (2nd Report, Session 2024–26, HL Paper 71), paras 103–11

111 Department for Science, Innovation and Technology, *Artificial Intelligence sector study 2024*, 3 September 2025

38. **Under existing law, copyright is engaged whenever the whole or a substantial part of a protected work is copied, including by storing it in digital form, subject only to specific statutory exceptions. We believe that the large-scale making and processing of digital copies of protected works for model training may therefore be characterised as reproduction, regardless of whether trained models retain human-readable copies or are capable of generating novel outputs. In our view, the lawfulness of using copyrighted content for AI training must be assessed under ordinary copyright principles and clearly defined exceptions. We do not accept the view that the copying and processing of protected works during training should be characterised as 'learning'.**

39. ***We therefore consider that the Government should rule out any reform of the Copyright, Designs and Patents Act that would remove the incentive to license copyrighted works for AI training, and should instead focus on strengthening licensing, transparency and enforcement within the existing framework.***

The impact of generative AI on the creative industries

Harms

40. Creative industry witnesses to our inquiry argued that, at present, copyrighted works are being accessed at multiple stages of the AI lifecycle—during large-scale training of foundation models, during fine-tuning, and, in some cases, at runtime to 'ground' systems through RAG—without consent, transparency or payment.[112] Anna Ganley, Chief Executive Officer of the Society of Authors, told us that this often involves mass scraping of music, books, images and journalism from the web, including from so-called "pirate libraries". In her view, this amounts to "industrial-scale" copying which, in other contexts, would be licensed and remunerated.[113] Isabelle Doran, Chief Executive Officer of the Association of Photographers, shared a statement from one of the organisation's members that sets out the economic impact starkly:

> "Generative AI undermines my ability to control the use of my work and therefore my IP income is reduced to 0%."[114]

41. A 2024 global study commissioned by the International Confederation of Societies of Authors and Composers predicts that generative AI will cause a €22 billion cumulative revenue loss for music and audiovisual creators by 2028 due to unlicensed content usage and market substitution. It estimated that widespread use of AI tools and surges in the AI-generated content market could put the revenues of 24% of music creators and 21% of audiovisual creators at risk.[115]

42. Witnesses to our inquiry similarly stressed that unlicensed use of protected works at the input stage of AI model development feeds directly into risks at

---

112 Q 2 (Tom Kiehl, Isabelle Doran, Anna Ganley, Owen Meredith), written evidence from Independent Society of Musicians (AIC0014) and supplementary written evidence from News Media Association (AIC0022)

113 Q 2 (Anna Ganley)

114 Q 2 (Isabelle Doran)

115 CISAC/PMP Strategy, *Study on the economic impact of Generative AI in the Music and Audiovisual industries*, November 2024, pp 71, 86

the output stage, since outputs may compete with original human-created works. Ms Doran told us that more than 58% of surveyed photographers had already lost work to generative AI.[116] Among authors, 86% reported decreased earnings as a result of AI advances, with 57% saying they no longer view their work as a sustainable career.[117] A recent report estimates that musicians face income cuts of 40% to 50% due to reduced commissions.[118] AI systems that can generate works 'in the style of' identifiable creators also threaten business models that depend on the distinctiveness and controlled reuse of their work, Ms Doran explained.[119]

43. Tom Kiehl, Chief Executive Officer of UK Music, drew attention to the proliferation of "inferior and low-quality" AI-generated content, or "AI slop", on algorithmically-driven streaming services.[120] Using its own AI-music detection tool, the French music streaming service Deezer revealed in September 2025 that 30,000 "fully AI-generated" tracks were being uploaded to the platform every day.[121] In January 2026, Deezer said that figure had doubled to 60,000 AI tracks a day, representing 39% of total daily uploads.[122] Since such content is often not clearly labelled, it can compete with music created by human artists for listener attention streams, diluting revenues for artists and rightsholders.[123] A case study reported by the ISM describes how an AI-generated project cloned the vocal style of the Welsh rock band Holding Absence and accumulated millions of streams on Spotify in a matter of months, including from listeners who believed "they were hearing human artists or Holding Absence themselves".[124]

44. We also heard concerns about the use of AI to create unauthorised 'digital replicas',[125] sometimes called 'deepfakes',[126] and synthetic performances.[127] A 2024 report by CREATe documented cases in which performers were not hired because AI systems were used instead, including voice actors being replaced in commercial work by synthetic voices. The report warned that AI-generated replicas could displace performers from large parts of the labour market, particularly lower-paid and routine roles.[128] Several witnesses emphasised that this trend of AI outputs substituting for, or devaluing, human work in a range of creative labour markets is particularly concerning given that these sectors are dominated by freelance workers and small businesses.[129]

---

116 Q 2 (Isabelle Doran)
117 Supplementary written evidence from Society of Authors (AIC0003)
118 Independent Society of Musicians, *Brave New World? Justice for creators in the age of GenAI*, 30 January 2026, p 11, hereafter referred to as *Brave New World?*
119 Q 2 (Isabelle Doran)
120 Q 2 (Tom Kiehl)
121 Deezer, *Deezer: 28% of all music delivered to streaming is now fully AI-generated*, 11 September 2025
122 Deezer, *Deezer confirms demonetization of up to 85% of AI-music streams due to fraud and moves to sell AI-detection Technology*, 29 January 2026
123 Q 2 (Tom Kiehl)
124 Independent Society of Musicians, *Brave New World?*, p 21
125 The Government defines digital replicas as "images, videos and audio recordings created by digital technology to realistically replicate an individual's voice or appearance." Intellectual Property Office, *Copyright and AI: Consultation*, p 30
126 According to Ofcom, "deepfakes are AI-generated audio-visual content that are deliberately designed to misrepresent someone or something. They pose a significant threat to online safety and can cause harm in myriad ways." Examples of harms include promoting financial scams or depicting people in non-consensual sexual imagery. Ofcom, *Deepfake Defences 2: The Attribution Toolkit*, 11 July 2025, p 3
127 Written evidence from British Equity Collecting Society (AIC0019)
128 CREATe Centre, *New Report : UK Audiovisual Performers' earnings and contracts*, 3 October 2024
129 Q 2 (Isabelle Doran, Tom Kiehl) and supplementary written evidence from Society of Authors (AIC0003)

45. For publishers, AI-generated summaries integrated into search results may compete by satisfying users' information requirements without them needing to click through to the publishers' original source material. Recent Ofcom research described a "Google Zero" effect, where "Google provides direct answers within its own ecosystem, leaving little incentive for users to visit publishers". It linked this to reduced visibility for publisher brands and declining referral traffic, which would pose a potential threat to advertising and subscription-based revenue models even where AI summaries are accurate and attributed.[130]

Barriers to enforcement

46. The legal test for copyright infringement is whether "the whole, or a substantial part of, a copyright protected work [has] been copied without permission or the use of an exception".[131] However, rightsholders told us that even where they suspect that their work may have been used to develop an AI model, there is currently no practical way of checking whether their material was in fact ingested.[132] This was primarily attributed to a lack of transparency from AI developers regarding the datasets or sources they have used for AI training.[133] We discuss this further in Chapter 3.

47. Enforcement is further complicated by the international nature of AI development. We heard that most of the early stages of AI development, including large-scale data collection, processing and model training, typically occur outside the UK.[134] Dr Alina Trapova, Lecturer in Intellectual Property Law at University College London, described this as an "elephant in the room". She noted that training often occurs in jurisdictions where copyright operates differently, particularly the United States,[135] and that UK courts may in some cases lack jurisdiction altogether.[136]

48. In principle, it is more straightforward to establish infringement where an AI system reproduces all or a substantial part of a protected work in its outputs. Recent technical research has demonstrated that, using specialist methods, it is possible in some cases to cause models to regurgitate large amounts of in-copyright text, including near-verbatim copies of entire books.[137] A recent ruling in Germany, currently under appeal, held that both such extracted outputs and the memorisation of those works encoded in a model's weights can constitute infringing copies of in-copyright training data.[138]

49. We heard that rightsholders who attempt to 'red team' models—probing these with targeted prompts, in this case with the aim of generating outputs

130 Ofcom, *The Era of Answer Engines*, 4 November 2025, p 36
131 Written evidence from Dr Hayleigh Bosher to the Communications and Digital Committee's inquiry on large language models (LLM0109)
132 Q 2 (Anna Ganley), written evidence from Copyright Licensing Agency (AIC0005) and written evidence from Music Publishers Association (AIC0017)
133 QQ 2 (Anna Ganley), 3–7 (Isabelle Doran), 4 (Tom Kiehl), 6 (Owen Meredith), 45 (Ed Newton- Rex), 51 (Serena Dederding), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from Copyright Licensing Agency (AIC0005) , written evidence from Independent Society of Musicians (AIC0014), written evidence from Music Publishers Association (AIC0017) and supplementary written evidence from Society of Authors (AIC0003)
134 Q 57 (Anthony Walker), 61 (Vinous Ali) and written evidence from HM Government (AIC0013)
135 See Table 2.
136 Supplementary written evidence from Dr Alina Trapova (AIC0021)
137 Ahmed Ahmed, A. Feder Cooper, Sanmi Koyejo, Percy Liang, 'Extracting books from production language models', (2026): https://doi.org/10.48550/arXiv.2601.02671
138 Reuters, *OpenAI used song lyrics in violation of copyright laws, German court says*, 11 November 2025

that would reveal the use of copyrighted material in the model's training[139]—find "the process takes time and does not always work".[140] In addition, technology sector witnesses emphasised that newer AI systems are explicitly designed to minimise this kind of direct recitation.[141]

50. A further challenge is that even where outputs clearly infringe copyright, these are difficult to detect in reality. Ms Doran cited estimates that there are now 34 million new synthetic images being created each day. She described the task of identifying a commercially significant infringement within this volume of material as "like finding a needle in a haystack", making it unrealistic for individual creators or small rightsholders.[142]

51. The perceived risks and costs of litigation act as an additional deterrent for rightsholders. The Copyright Licensing Agency (CLA) told us that enforcement in the UK is already prohibitively expensive for many creators and rightsholders, given the high cost of legal proceedings and the lack of statutory awards for damages.[143] Ms Doran raised similar points. She also warned of rightsholders' concerns that an adverse legal judgment could entrench large-scale infringements.[144]

52. While acknowledging the practical barriers to enforcing their rights through the courts, witnesses underlined the importance of compensation for past, unlicensed uses of creative work for AI training. Ms Ganley told us that members of the Society of Authors "are looking for compensation for works that have already been scraped", stressing that "compensation retrospectively is one part of it", alongside regulation for the future.[145] The Authors' Licensing and Collecting Society reported that 92% of surveyed members wanted compensation for any historic use of their work for AI training.[146] We note that the Science, Innovation and Technology Committee has described as "inevitable" both "the agreement of a financial settlement for past infringements by AI developers" and the negotiation of licensing frameworks for future uses.[147] Other stakeholders have suggested "levies" on AI firms to fund cultural or public-interest schemes,[148] and AI or data "dividends" that would distribute a share of AI-generated value to creators or the wider public.[149]

---

139 IBM, *What is red teaming for generative AI?*, 11 April 2024

140 Q 45 (Ed Newton-Rex)

141 Q 109 (Roxanne Carter) and written evidence from Microsoft (AIC0012)

142 Q 2 (Isabelle Doran)

143 Written evidence from Copyright Licensing Agency (AIC0005). Under UK law, remedies for copyright infringement are primarily compensatory damages or an account of profits, with the court also having a discretion to award "additional damages" in cases such as flagrant infringement. These remedies are assessed case-by-case by the court and there is currently no statutory damages regime setting fixed-sum awards per work infringed irrespective of proven loss, unlike in other jurisdictions such as the US. Slaughter and May, *Statutory damages for IP infringement: 'It's a wonderful world'?*, 22 July 2021

144 Q 2 (Isabelle Doran)

145 Q 2 (Anna Ganley)

146 Authors' Licensing and Collecting Society, *UK AI Regulation: principles and the way forward*, December 2024

147 Science, Innovation and Technology Committee, *Governance of artificial intelligence (AI)* (3rd Report, Session 2023–24, HC Paper 38), para 186

148 Professor Martin Senftleben of the University of Amsterdam proposes that such levies should be imposed at the output level on commercial providers of generative AI systems rather than the training stage because AI systems compete with human works in the marketplace: Martin Senftleben, 'Generative AI and Author Remuneration', *SSRN*, January 2023: https://doi.org/10.2139/ssrn.4478370

149 Nicholas Vincent and Brent Hecht, *Sharing the Winnings of AI with Data Dividends: Challenges with "Meritocratic" Data Valuation*, 2023

Potential opportunities

53. Alongside setting out the risks and barriers described above, our evidence also indicated that generative AI may offer opportunities for the creative industries.[150] First and foremost, witnesses identified that licensing content for AI use could open up new revenue streams for rightsholders.[151] This would, we heard, be contingent on having effective transparency arrangements and technical controls in place. We explore these areas in more detail in later chapters.

54. Others emphasised that many parts of the creative industries are already using AI and other technologies to support productivity, innovation and growth.[152] The Creative Industries Policy and Evidence Centre (Creative PEC) noted that the sector has benefited from technological innovation alongside a robust copyright framework and highlighted the rise of UK 'createch' firms—businesses that integrate creative practice with advanced technologies—which now constitute a growing part of the creative economy.[153] As detailed in our 2025 report on scaleups, these businesses have significant economic potential.[154] It is promising to see that such firms have been explicitly recognised for targeted support in the Government's *Creative Industries Sector Plan*.[155]

55. The Government told us that nearly 40% of createch business had already adopted AI technologies by September 2024, and that createch businesses are expected to generate £18 billion in GVA over the next decade. It stated that "AI will therefore underpin growth and development across the creative sectors, reinforcing their role as a driver of innovation, productivity, and economic expansion."[156]

56. **The UK creative industries are an economic powerhouse with huge growth potential, and a vital part of the UK's soft power. Their success has been underpinned by what witnesses described as an internationally-respected "gold-standard" copyright framework. While generative AI may present opportunities for creative innovation and further growth where appropriate safeguards are in place, it also poses substantial material risks to the work and livelihoods of individual creators. The impact of these is already being felt.**

57. **We recognise that, in relation to AI-generated outputs that directly reproduce or 'regurgitate' protected works, existing copyright doctrines may provide a route in principle to challenge infringement. In practice, however, it is often prohibitively difficult for rightsholders**

150 QQ 2 (Tom Kiehl), 68 (Antony Walker), 125 (Guy Gadney), written evidence from CREATe Centre, University of Glasgow (AIC0007), written evidence from Creative Industries Policy and Evidence Centre (AIC0008), written evidence from Meta (AIC0009), written evidence from Google (AIC0010) and written evidence from Music Publishers Association (AIC0017)

151 QQ 3 (Anna Ganley), 8 (Tom Kiehl), 36 (Ed Newton-Rex), 38 (Reema Selhi, Serena Dederding), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from Copyright Licensing Agency (AIC0005), written evidence from techUK (AIC0011), written evidence from Independent Society of Musicians (AIC0014) and written evidence from Music Publishers Association (AIC0017)

152 QQ 2 (Tom Kiehl), 65 (Vinous Ali), written evidence from Meta (AIC0009), written evidence from Google (AIC0010) and written evidence from Music Publishers Association (AIC0017)

153 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)

154 Communications and Digital Committee, *AI and creative technology scaleups: less talk, more action*, (2nd Report, Session 2024–26, HL Paper 71), paras 142–53

155 HM Government, *Creative Industries Sector Plan*, 23 June 2025, p 9

156 Written evidence from HM Government (AIC0013)

to detect, evidence and enforce such claims. Despite this, the courts may be the only viable means of recovering damages for technology companies' past use of content for training purposes.

58. **AI has the potential to contribute to the UK's future economic growth and there could be a legitimate interest in enabling more AI training to take place on UK soil. However, in 2024 the UK AI sector contributed around £12 billion to the UK economy, compared with £124 billion in gross value added from the creative industries in 2023. The evidence available to date has not convincingly demonstrated that the existing copyright framework is the primary constraint on AI investment, or that weakening it through new or broader exceptions for commercial AI training would, in itself, significantly expand the sector.**

59. **At the same time, the creative industries are already experiencing tangible harms from the current deployment of generative AI. Given their proven economic and cultural value, and the lack of robust evidence on the likely benefits of copyright reform to the UK AI sector, we do not consider that the case has been made for introducing new exceptions for AI training or for undertaking wider reform of UK copyright law. We welcome the Government's forthcoming economic impact assessment as an essential basis for any future, evidence-led decisions on whether changes to copyright law are justified. It is urgent that a regime is now created to safeguard creators' livelihoods going forward, while harnessing the potential of AI for creativity and economic growth.**

The Government's approach and the need for clarity

60. We have previously stressed the importance of the Government swiftly reaching a definitive position on AI and copyright.[157] These issues have been under consideration since the Intellectual Property Office launched a consultation on the relationship between intellectual property and AI in 2021.[158] It set out its conclusions in 2022, proposing a new copyright and database right exception allowing TDM for any purpose.[159] The then Government subsequently confirmed that it would not pursue a "broad copyright exception". It committed to setting up a working group to develop a voluntary code of practice on the training of AI models using copyrighted material by summer 2023.[160] However, that group was unable to reach agreement, and the planned code was not taken forward.[161]

61. We are now in 2026, and a year on from the current Government's consultation. It appears, however, that the Government is still some way from setting out a clear policy direction. Giving evidence to our inquiry, the Rt Hon Liz Kendall MP, Secretary of State for Science, Innovation and Technology, and the Rt Hon Lisa Nandy MP, Secretary of State for Culture,

157 See, for example, Communications and Digital Committee, *Large language models and generative AI* (1st Report, Session 2023–24, HL Paper 54), paras 245–49, and *AI and creative technology scaleups: less talk, more action* (2nd Report, Session 2024–25, HL Paper 71), para 178.

158 Intellectual Property Office, *Artificial Intelligence and IP: copyright and patents*, 29 October 2021

159 Intellectual Property Office, *Artificial Intelligence and Intellectual Property: copyright and patents: Government response to consultation*, 28 June 2022

160 Correspondence, *Minister for Media, Tourism and Creative Industries to the Chair of the Communications and Digital Committee*, 18 April 2023

161 Financial Times, *UK shelves proposed AI copyright code in blow to creative industries*, 5 February 2025 [paywall]

Media and Sport, implied that the Government would not be setting out its position in March, when updates on this area are due under the Data (Use and Access) Act.[162]

62. The Secretary of State for Science, Innovation and Technology told us that this is because the Government is having a "reset moment" and "genuinely consulting to try and find a way forward".[163] The Secretary of State for Culture, Media and Sport explained:

> "If we rush into this and get it wrong, we could make a mess. We are not going to rush into it; we are going to take the time to work through it with the working groups".

She assured us, however, that "we appreciate the urgency … and want to move as quickly as possible."[164] The Secretary of State for Science, Innovation and Technology acknowledged that any eventual settlement would need to strike a balance between competing interests, but that "not everybody will get everything—that is the nature of this".[165]

63. A minority of witnesses supported the Government's decision to proceed cautiously rather than legislate at speed. Creative PEC noted that the rapid evolution of generative AI since 2019 creates a "real risk" that moving too quickly to reform copyright law could lead to unintended and counterproductive outcomes.[166] Dr Trapova likewise advocated an evidence-based approach, warning against "regulating for the sake of regulating" in a highly emotive area and urging policymakers to "slow down".[167]

64. Otherwise, there was consensus that the Government's deliberation on this issue must be quickly brought to a close. The British Copyright Council warned that the "protracted consultation process" and continued suggestion of a possible commercial TDM exception had "generated uncertainty" that undermines the Government's stated ambition to promote licensing.[168] The Creative Rights in AI Coalition similarly argued that a "drawn-out consultation process" and "wrongly questioning the clarity provided by UK copyright law … has undoubtedly disincentivised AI developers from striking deals with UK rightsholders".[169] Mr Newton-Rex raised similar points.[170]

65. Reema Selhi, Head of Policy and International at the Design and Artists Copyright Society (DACS), reported that when DACS approached image-generating AI companies about entering into copyright licences, many "basically said, 'Well, let's just wait and see where the Government consultation will go on copyright and AI'."[171] Mr Meredith went further, stating:

---

162 QQ 136–38 (Secretary of State for Culture, Media and Sport)
163 Q 136 (Secretary of State for Science, Innovation and Technology)
164 Q 138 (Secretary of State for Culture, Media and Sport)
165 Q 149 (Secretary of State for Science, Innovation and Technology)
166 World Intellectual Property Organization, *Artificial Intelligence and Intellectual Property: An Economic Perspective*, 2024, as cited in: written evidence from Creative Industries Policy and Evidence Centre (AIC0008).
167 Q 19 (Dr Alina Trapova)
168 Written evidence from British Copyright Council (AIC0006)
169 Written evidence from Creative Rights in AI Coalition (AIC0001)
170 Q 36 (Ed Newton-Rex)
171 Q 38 (Reema Selhi)

> "The Government in some ways has been complicit in creating uncertainty in the market through the way it has approached its policy agenda in this area".

This, he argued, has "allowed developer firms to exploit that uncertainty to their own benefit."[172]

66. Several creative industry witnesses highlighted the recent announcement by the Australian Government that it will not introduce a new TDM exception, and urged the UK Government to make an equivalent commitment.[173] According to the Creative Rights in AI Coalition,

> "this demonstrates that a far more productive conversation on avenues to support the creative industries and AI firms can be opened up if a damaging change in copyright law is taken off the table."[174]

Serena Dedering, General Counsel and Company Secretary at the CLA, Mr Newton-Rex and Ms Selhi all argued that ruling out a new commercial TDM exception would be the most effective step that the Government could take at this stage to help a licensing market to flourish in the UK.[175]

Box 2: The Australian Government's statement on AI and copyright

In October 2025, the Australian Government announced that although it is investigating possible updates to Australian copyright law in relation to AI training, this will not include the introduction of a broad TDM exception, stating that this approach would provide "certainty to Australian creators".

The Australian Government has convened its Copyright and AI Reference Group to examine whether "a new paid collective licensing framework under the Copyright Act should be established for AI"; whether clarifications are needed on how copyright law applies to AI-generated material; and "avenues for less costly enforcement", such as a "potential new small claims forum" aimed at making it easier for rightsholders to enforce existing rights.

Source: Attorney-General's Department, Albanese Government to ensure Australia is prepared for future copyright challenges emerging from AI, 26 October 2025

67. Meta also argued that "legal clarity and certainty" must underpin any approach to copyright if the Government is to "enhance UK competitiveness and influence in AI", even if they—along with other technology companies—disagreed with rightsholders about the substantive direction of reform.[176]

68. **The application of copyright law to the development and operation of AI models is a technically complex and contested area. The Government is therefore right to seek evidence and engage extensively before reaching a decision, and we very much welcome that it has "reset" its previous approach. However, prolonged uncertainty and mixed public messaging on AI and copyright since this Government**

172 Q 2 (Owen Meredith)
173 QQ 50, 55 (Reema Selhi), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from Copyright Licensing Agency (AIC0005), written evidence from British Copyright Council (AIC0006), supplementary written evidence from Society of Authors (AIC0003) and supplementary written evidence from News Media Association (AIC0022)
174 Written evidence from Creative Rights in AI Coalition (AIC0001)
175 Q 50 (Serena Dederding, Ed Newton-Rex, Reema Selhi)
176 Written evidence from Meta (AIC0009)

came into office have undermined trust and stalled both licensing and investment.

69. **The Government must accept that there is no solution that will satisfy all parties and that delaying in the hope of finding one risks exacerbating these issues further. Committing to a clear, stable policy framework, rooted in the UK's longstanding copyright principles and in the interests of its domestic creative and AI sectors, is increasingly urgent.**

70. ***The Government should reach and publish a final, evidence-based decision on its approach to AI and copyright in the next 12 months. This decision should make clear that strong copyright protection and fair licensing for UK rightsholders are the default, and that policy choices will prioritise the long-term health of the UK's creative industries and domestic AI sector, rather than the interests of large multinational technology firms. In its March update published under the Data (Use and Access) Act, the Government should set out the concrete steps and timetable it will follow to reach this decision, and detail any further evidence, analysis, or engagement it requires.***

71. ***In the meantime, we urge the Government to issue a clear public statement setting out an expectation that commercial AI developers operating in the UK should obtain appropriate licences when using copyrighted works to train generative AI models. That statement should signal the Government's support for a licensing-first approach as the baseline for the UK market, and should be published at the earliest opportunity.***

Gaps in protection for style imitation and digital identity

72. While our evidence did not present a clear case for reform of the UK's copyright framework to benefit the UK's AI sector, it did identify one area where legislative change may be necessary. We heard that one of the most acute harms arising from generative AI outputs is outputs 'in the style of' an artist. These can imitate their recognisable style, voice or personality in a way that displaces commissions or devalues the distinctive appeal of their work without reproducing a substantial part of any particular work.[177]

73. Copyright protects specific expressions, not general ideas, themes or style.[178] As Dr Guadamuz noted, this approach is "foundational to copyright systems worldwide". He explained: "Protecting style under copyright would come dangerously close to granting monopolies over artistic techniques and conventions, which would chill the very creativity that copyright is designed to promote."[179] AI systems can therefore generate convincing stylistic imitations without necessarily engaging copyright in any individual underlying work, leaving a structural gap for creators whose economic position depends on the distinctiveness of their style.

74. We also heard that copyright offers only partial and often indirect protection against deepfakes and other digital replicas that reuse a person's image or voice. A report by the ISM notes that "performers are among the most severely affected by the rapid spread of digital replicas, which allow voices,

177 Written evidence from Independent Society of Musicians (AIC0014)
178 Written evidence from Dr Hayleigh Bosher (AIC0025)
179 Written evidence from Dr Andres Guadamuz (AIC0026)

faces and performances to be captured once and reused indefinitely without further consent, control or payment".[180] Dr Bosher highlighted that, in many cases, the person depicted or recorded is not the copyright holder; the rights often lie with the photographer, producer or commissioning entity, meaning victims of deepfakes cannot themselves bring a copyright claim unless they also own the relevant rights. Even where they do, Dr Bosher explained that they must also overcome the test for infringement, which can prove challenging.[181]

75. Dr Guadamuz observed that the UK has no statutory "voice likeness right", and the human voice itself is not protected by copyright. Performers' rights under Part II of the CDPA offer some safeguards against unauthorised recording and exploitation, but there is no right of adaptation for performances, so once a recording has been made with consent it can be difficult to prevent its use for voice cloning. The tort of passing off may assist only those with established goodwill, limiting its reach to celebrities and well-known performers.[182]

76. Taken together, these features produce a major gap in the UK's legal framework. As the ISM's report put it: "while copyright protects creative works, it does not fully protect the person who creates them."[183] Google told us that the proliferation of unauthorised digital replicas was among the "most pressing concerns" relating to intellectual property and generative AI.[184]

77. Several witnesses therefore advocated for the introduction of a UK 'personality right'[185] that would give individuals greater control over how their likeness or voice is used.[186] Ms Doran suggested that such a right could sit alongside trade mark and passing off protections,[187] rather than within copyright law.[188]

78. The Government's AI and copyright consultation acknowledged creative industry stakeholders' interest in the introduction of personality rights. It welcomed input from respondents on the adequacy of the current legal framework and stated that "there will be opportunities to engage with us

---

180 Independent Society of Musicians, *Brave New World?*, p 22
181 Written evidence from Dr Hayleigh Bosher (AIC0025)
182 Written evidence from Dr Guadamuz (AIC0026)
183 Independent Society of Musicians, *Brave New World?*, p 24
184 Written evidence from Google (AIC0010)
185 "Personality rights, also known as 'image rights' or 'publicity rights', refer to the right of individuals to control the use of their name, image, likeness, or other personal attributes such as their nickname and signature." The UK currently has no specific statutory protections of this kind. CMS Law-Now, *Personality rights in the UK—Possible reform on the horizon to address AI-related challenges*, 28 January 2025
186 QQ 15 (Isabelle Doran), 30 (Prof Eleanora Rosati), 54 (Ed Newton-Rex, Reema Selhi), written evidence from Independent Society of Musicians (AIC0014), written evidence from British Equity Collecting Society (AIC0019), written evidence from Dr Hayleigh Bosher (AIC0025) and written evidence from Dr Andres Guadamuz (AIC0026)
187 Since under current UK law there is no standalone personality or image right, individuals instead rely on a patchwork of causes of action, notably registered and unregistered trademarks and the common-law tort of passing off (which protect commercial goodwill and prevent false endorsement), together with copyright, performers' rights, privacy and data protection law. See written evidence from Dr Hayleigh Bosher to the Communications and Digital Committee's inquiry on large language models (LLM0109); Intellectual Property Office, *Copyright and AI: Consultation*; and Martin Kretschmer et al, 'Copyright and AI in the UK: Opting-In or Opting-Out?', *GRUR International*, vol 74, Issue 11, (November 2025), pp 1055–70: https://doi.org/10.1093/grurint/ikaf093.
188 Q 15 (Isabelle Doran)

on this topic in the future."[189] Potential new personality rights have been a focus of the recent Government-led technical working group on "wider protections for creators".[190]

79. The British Equity Collecting Society (BECS), a collective management organisation for audiovisual performers, pointed to the US Copyright Office's recognition of the "urgent need for legislation to prevent harm resulting from unauthorised digital replicas". Dr Bosher noted that, while many US states recognise publicity or image rights protecting the commercial use of a person's name, likeness or voice, these typically focus on celebrities and commercial advertising, leaving non-commercial harms and non-celebrity victims less well covered. The US Copyright Office has therefore advocated for new federal legislation extending protection to all individuals, regardless of the commercial value of their identity.[191] BECS identified this emerging US approach as a useful model, and said that any UK framework should likewise allow individuals to license and monetise their digital replica rights, but not to assign them outright. [192]

80. BECS also drew attention to the Danish government's bill to amend its Copyright Act to protect individuals against AI-generated deepfakes and unauthorised digital replicas of their likenesses, including face, voice and other identifying characteristics.[193] In addition, BECS called for the introduction of unwaivable moral rights for audiovisual performers, ideally through the ratification of the Beijing Treaty on Audiovisual Performances.[194]

81. Academic commentators have emphasised that the design of any new image or personality right would require careful consideration, noting that overly broad rights could have unintended effects on freedom of expression, for example by inhibiting biographical and historical works, or unduly constraining news reporting. They stressed that any new statutory right

189 Intellectual Property Office, *Copyright and AI: Consultation*, p 31

190 HM Government, *Copyright and AI technical working group: roundtables*, 15 December 2025

191 Written evidence from Dr Hayleigh Bosher (AIC0025)

192 Written evidence from British Equity Collecting Society (AIC0019)

193 The proposed amendments to the Danish Copyright Act are composed of two parts: a general protection against realistic digitally generated imitations of a natural person's personal, physical characteristics (such as the face, body, or voice) being made available to the public without consent; and a specific protection for performing artists against realistic, digitally generated imitations of their artistic performances or achievement. Communication from the Commission: Notification Detail—Act amending the Copyright Act (Introduction of performance protection and protection against digitally generated imitations, etc.), 2025/0654/DK (Denmark), 31 October 2025

194 The Beijing Treaty on Audiovisual Performances is a treaty administered by the World Intellectual Property Organization (WIPO), adopted in 2012, that extends performers' economic and moral rights in audiovisual performances, including films, television programmes and other recordings. WIPO, *The Beijing Treaty*. It gives performers, among other things, rights to be identified and to object to prejudicial distortions of their recorded performances. The UK signed the Treaty in 2013 but has not yet ratified it, so these additional protections are not currently fully reflected in UK law. WIPO, *WIPO-Administered Treaties*. In our report *At risk: our creative future*, we recommended that the Government ratify the Beijing Treaty "at the earliest opportunity". Communications and Digital Committee, *At risk: our creative future* (2nd Report, Session 2022–23, HL Paper 125), paras 40–42. Since then, the Government has stated that it is committed to implementing and ratifying the treaty and has consulted on options for implementation. Culture, Media and Sport Committee, *Connected tech: AI and creative technology: Government Response to the Committee's Eleventh Report of Session 2022–23* (HC Paper 441), p 6; Intellectual Property Office, *Consultation on the options for implementing the Beijing Treaty on Audiovisual Performances*, 7 February 2024. However, as of May 2025, it was still "considering options for ratifying the Beijing Treaty": HC Written Answer, 47564, 1 May 2025

would need to be tightly defined and limited in scope.[195] We note that the Danish model also seeks to address these concerns through a carve-out for imitations that are mainly expressions of caricature, satire, parody, pastiche, or criticism of power or society, unless they constitute misinformation likely to cause serious harm to others' rights or essential interests.[196]

82. Dr Trapova said that new protections from the harms posed by deepfakes should not be framed as providing "a licensing opportunity or another stream of revenue" for individuals. In her view, "this area deserves legislation of its own", rather than being tackled through an expansion of copyright.[197] In its report, the ISM proposed that:

> "New statutory personality rights, or reforms to existing performer and image rights, could sit alongside the copyright and performance rights framework and be supported by related areas of law, including data protection and privacy law, consumer protection and misrepresentation law, contract law or competition law."[198]

83. **The absence in UK law of a robust personality right or specific protection for digital likeness means that creators and performers lack an adequate basis to challenge harmful outputs that imitate their distinctive style, voice or persona without reproducing a specific underlying work.**

84. ***The Government should introduce protections against unauthorised digital replicas and 'in the style of' uses. Any new framework should give creators and performers clear, enforceable control over the commercial exploitation of their identity, while appropriately safeguarding freedom of expression and other legitimate uses.***

195 See Martin Kretschmer et al, 'Copyright and AI in the UK: Opting-In or Opting-Out?', *GRUR International*, vol 74, Issue 11, (November 2025), pp 1055–70: https://doi.org/10.1093/grurint/ikaf093. Google raised similar concerns and cited, along with the British Equity and Collecting Society, the United States Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES) Act of 2025 as an alternative model. Written evidence from Google (AIC0010) and British Equity and Collecting Society (AIC0019)

196 EM Law, *Denmark Proposes to Protect Individuals From AI Deepfakes by Making Changes to its Copyright Laws*, 23 July 2025; European Parliamentary Research Service, *The Danish approach to copyright and deepfakes: A model for the EU,* PE 782.611, January 2026

197 Q 31 (Dr Alina Trapova)

198 Independent Society of Musicians, *Brave New World?*, p 24

CHAPTER 3: TRANSPARENCY

85. We use 'transparency' here to refer to disclosures by AI developers regarding the material they have used to train their models. In the UK, developers are not currently required to publish this information. As the Government's AI and copyright consultation noted:

> "Some developers use works from databases which are open to the public and can be scrutinised by third parties. But others do not disclose the source of the works they use in training, or only disclose limited information about them."[199]

86. Our evidence made clear that increased transparency is the key battleground in the development of an effective, licensing-first ecosystem. As such, it demonstrated that any disclosure requirements must be designed carefully to ensure they are effective and proportionate. This chapter discusses these points in more detail and sets out key considerations for future UK transparency requirements.

The case for increased transparency

87. As noted in Chapter 2, rightsholders told us that a lack of transparency from AI developers weakens their protections under copyright law by making it harder to identify infringement of their material.[200] In addition, they cannot monitor whether any rights-reservation signals they may have used to prevent their content being accessed for AI training[201] have been respected.

88. Witnesses were clear that the lack of any UK transparency requirements means there is little incentive for developers to license copyrighted content that they may be able to access without remuneration. As such, increased transparency was seen as central to unlocking a licensing market for AI training.[202] Ed Newton-Rex, Chief Executive Officer of Fairly Trained, described requiring transparency in training data as "the single most important and impactful thing" the Government could do to support licensing.[203]

89. The Creative Rights in AI Coalition argued that this "reduction in informational asymmetries will also help rightsholders to negotiate for the true value of their works" in licensing arrangements.[204] Reema Selhi, Head of Policy and International at the Design and Artists Copyright Society (DACS), made a similar point, noting that rightsholders "need some information in order to know how to license, what we are licensing and what the value of that licence is."[205]

90. The Alliance for Intellectual Property told us that transparency is also "important for wider public interest reasons, including privacy."[206] In a similar vein, Owen Meredith, Chief Executive Officer of the News Media Association, commented:

---

199 Intellectual Property Office, *Copyright and AI: Consultation*, pp 19–20
200 See para 46.
201 See Chapter 4.
202 Q 50 (Ed Newton-Rex), written evidence from Creative Rights in AI Coalition (AIC0001) and written evidence from Copyright Licensing Agency (AIC0005)
203 Q 51 (Ed Newton-Rex)
204 Written evidence from Creative Rights in AI Coalition (AIC0001)
205 Q 44 (Reema Selhi)
206 Written evidence from Alliance for Intellectual Property (AIC0002)

> "It is clear that transparency will be fundamental in tackling a range of AI policy concerns—ensuring compliance with data protection law, safety testing, preventing bias and prejudice, and ensuring that illegal content such as CSAM is not used in training will all require transparency from AI developers. Creative industry calls for transparency should be seen in this context, rather than as an unfair trade-off against a severe degradation in fundamental rights".[207]

91. Serena Dedering, General Counsel and Company Secretary at the Copyright Licensing Agency, argued that transparency "works for the benefit of AI developers as well", since it "can be a huge lever in general users' understanding of what data has been used—for example, to train that model—and its ethical, legal, compliance credentials." This could, she suggested, increase public trust in AI systems and "lead to greater adoption and scalability".[208]

92. Tom Kiehl, Chief Executive Officer of UK Music, told us that there was an opportunity for the UK to lead internationally by developing "a gold standard around transparency".[209] Mr Newton-Rex likewise called for the UK to "lead the discussion on the international stage."[210] Ms Selhi felt that there was a clear role for Government in helping creative industry and AI sector stakeholders to "develop those transparency models together".[211] We were encouraged to hear the Rt Hon Lisa Nandy MP, Secretary of State for Culture, Media and Sport, echo this language:

> "There is a strong role for Government in introducing requirements around things like transparency, which industry cannot do itself … We have committed to this as a Government; we see that as part of our proper role."[212]

Considerations for transparency requirements

Granular transparency

93. We heard consistently from rightsholders that for transparency reporting to be meaningful, it must be sufficiently detailed; the high-level summaries required under the EU AI Act were felt to be inadequate.[213] According to the Creative Rights in AI Coalition, these do not provide "any granularity that would be useful to rightsholders in determining whether and how their copyright works have been used".[214]

---

207 Supplementary written evidence from News Media Association (AIC0022)
208 Q 49 (Serena Dederding)
209 Q 4 (Tom Kiehl)
210 Q 51 (Ed Newton-Rex)
211 Q 44 (Reema Selhi)
212 Q 142 (Secretary of State for Culture, Media and Sport)
213 Q 4 (Tom Kiehl, Owen Meredith) and written evidence from Alliance for Intellectual Property (AIC0002)
214 Written evidence from Creative Rights in AI Coalition (AIC0001)

Box 3: Transparency requirements under the EU's AI Act

The EU's AI Act builds on the 2019 Directive on Copyright in the Digital Single Market (CDSMD),[215] which introduced a two-tier TDM framework. The Act requires providers of general-purpose AI models placed on the EU market to adopt a policy to comply with EU copyright and related rights, "in particular" by identifying and respecting TDM opt-outs under the CDSMD through "state-of-the-art technologies", regardless of where the training took place.[216]

The Act further requires providers to make publicly available a "sufficiently detailed summary" of their training content, using a standardised template issued by the EU AI Office.[217] According to the EU Commission, the summary is intended to "increase transparency about the model's training data and assist parties with legitimate interests, such as copyright holders, in exercising their rights under Union law."

Guidance states that the summary must be "published on the provider's official website in a clearly visible and accessible manner", and "made publicly available alongside the model across all its public distribution channels, such as online platforms." This is a mandatory requirement; from 2 August 2026, non-compliance may result in enforcement action being taken.[218]

Alongside this, model providers can choose to sign the EU's General-Purpose AI Code of Practice, "a voluntary tool … designed to help industry comply with the AI Act's obligations".[219] The transparency chapter of the code includes a "model documentation form" that requests details of a model's training methodology; training data type (for example, text, images, video); data provenance (for example whether the data was obtained via web crawling or publicly available datasets, or includes content obtained under licence); and number of data points.[220] The recipients of this information are "downstream providers, the AI Office or national competent authorities". Recipients "are obliged to respect the confidentiality of the information obtained, in particular intellectual property rights and confidential business information or trade secrets".[221]

94. Creative industry stakeholders suggested that AI developers should disclose, at individual work-level, what content they have accessed, and how; what specific works were used in AI training and fine-tuning, and for what purpose; and details of how works are stored, processed and retained in AI models.[222] The Copyright Licensing Agency and the Music Publishers' Association emphasised the need to record metadata attached to any works used.[223] We also heard calls for companies to report on how specific training

215 See Table 2.
216 European Commission, *AI Act Explorer: Article 53—Obligations for providers of general-purpose AI models*
217 *Ibid.*
218 This could include fines of up to 3% of the provider's annual total worldwide turnover in the preceding financial year, or €15,000,000, whichever is higher. European Commission, *Template for general-purpose AI model providers to summarise their training content*, 31 July 2025
219 European Commission, *The General-Purpose AI Code of Practice*, 2 February 2026
220 'Model Documentation Form', available at European Commission, *The General-Purpose AI Code of Practice*, 2 February 2026
221 'Code of Practice for General-Purpose AI Models Transparency Chapter', available at: European Commission, *The General-Purpose AI Code of Practice*, 2 February 2026
222 QQ 3 (Isabelle Doran), 4 (Owen Meredith), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from Alliance for Intellectual Property (AIC0002) and written evidence from Independent Society of Musicians (AIC0014)
223 Written evidence from Copyright Licensing Agency (AIC0005) and Music Publishers Association (AIC0017). We discuss the challenge of ensuring that metadata associated with a particular work is preserved in Chapter 4, as well as potential technical solutions.

data shapes model outputs.[224] The Alliance for Intellectual Property advised that transparency measures "are likely to require supply chain practices which are specific to sub-sectors of IP markets and best arranged and agreed in those subsectors".[225]

95. The British Copyright Council proposed duties building on existing CDPA provisions that would "enforce the publication of a transparency usage report in relation to the use of copyright works". These reports should, it argued, "enable rightsholders to identify works to be used and how access is anticipated". As such, they would:

> "support enforcement actions which currently impose an impossible evidential burden on rightsholders … The regulation could provide that if an AI developer fails to provide a Transparency Usage Report, it will be subject of a rebuttable legal presumption of use, thus reversing the burden of proof. This means that the AI developer will have to prove that content has not been used."[226]

96. In response to the Government's consultation, the collective management organisations DACS and PICSEL commissioned workshops and a research report on AI and copyright. It set out that transparency must be built around three pillars:

- Legibility: Transparency measures introduced should be understandable without legal or technical expertise so individual creators can understand what information they are entitled to receive and how to obtain it.
- Auditability: Disclosures must be detailed enough for rightsholders or intermediaries to determine whether their works have been used and whether violation of copyright/IP rights has occurred.
- Impact: Information must enable rightsholders to seek redress (for example via a regulator or courts) where misuse is identified.[227]

97. DACS, PICSEL and the think tank Convocation Research and Design Labs also co-developed a proposed 10-point transparency model that AI developers could follow to achieve "meaningful" transparency. This proposes that developers should publish:

(1) sources and owners of each dataset, including those used to generate synthetic data, with URLs and descriptions for publicly available online sources;

(2) work-level information for data ingested (such as image size, metadata and labels);

(3) the timeframe of data collection;

(4) the location of the dataset;

---

224 Q 3 (Isabelle Doran) and written evidence from Independent Society of Musicians (AIC0014)
225 Written evidence from Alliance for Intellectual Property (AIC0002)
226 Written evidence from British Copyright Council (AIC0006)
227 Convocation Research and Design Labs, DACS and PICSEL, *Considering Creatives, Copyright and AI: A Response to the UK Government Consultation*, 8 April 2025, p 11

(5) the specific legal basis on which each dataset was ingested and processed;

(6) the weighting or percentage contribution of each data source to the overall training or fine-tuning;

(7) the region in which training took place;

(8) a description of the measures taken to ensure compliance with UK copyright law, "including measures taken to comply with any proposed opt-outs including those in other jurisdictions";

(9) details of any third-party providers of data; and

(10) the types of automated tools used to acquire content (for example crawlers), including those run by third parties.[228]

98. Technology sector witnesses raised several issues about disclosing data at this level of granularity. First, developers stressed the administrative burden of such reporting. Roxanne Carter, Global IP Lead for the Government Affairs and Policy team at Google, told us that the scale of such disclosures would be "too vast, even for Google". She explained: "If you are asking us to provide, URL by URL, a granular list … you are asking us to summarise the internet."[229] Meta expressed similar concerns: "Proposals that require highly granular disclosure fail to account for the massive scale of data involved in training AI models and the technical realities of this process."[230] Witnesses explained that since copyright is an 'unregistered right' in the UK, there is no register or repository of protected works that developers can reference to verify ownership, which creates additional complexity,[231] particularly when combined with "the poor quality of rights metadata".[232]

99. The trade body techUK also noted that requirements for granular transparency reporting "would likely result in high costs, both as a compliance cost but also a barrier to deploying new technologies".[233] Matthew Sinclair, Senior Director at the Computer and Communications Industry Association, agreed that it was not clear how these could be delivered "at a cost that makes it possible for the broad swathe of AI developers".[234] Emphasising the challenge for smaller companies, Vinous Ali, Deputy Executive Director at the Startup Coalition, told us that this kind of transparency reporting "requires a granularity that is impossible for startups to meet."[235]

100. techUK also acknowledged, however, a minority view among its members that "well-designed transparency requirements that align with existing practices may not result in disproportionate costs".[236] Mr Newton-Rex argued that "there is no extra process" involved in compiling this information, since AI

228 *Ibid.*, p 12
229 Q 117 (Roxanne Carter)
230 Written evidence from Meta (AIC0009)
231 QQ 27 (Dr Alina Trapova), 60 (Matthew Sinclair), 71 (Vinous Ali) and written evidence from Google (AIC0010)
232 Supplementary written evidence from Google (AIC0024). The topic of rights metadata is discussed in Chapter 4.
233 Written evidence from techUK (AIC0011)
234 Q 76 (Matthew Sinclair)
235 Q 71 (Vinous Ali)
236 Written evidence from techUK (AIC0011)

companies would typically keep detailed records of their training data.[237] Dr Andres Guadamuz, Reader in Intellectual Property Law at the University of Sussex, commented that while disclosing "every specific web page or asset used in training" would be "technically demanding", "it is technically feasible for developers to maintain records of the data sources used".[238]

101. Dr Guadamuz also noted that "publishing the full list" of data sources would raise "practical and commercial sensitivity concerns".[239] Similarly, OpenAI told us:

> "Requiring more detailed reporting, such as individual works or web URLs, would raise significant concerns related to model developers' trade secrets. AI developers invest heavily in refining their data and training processes, and publicising this knowledge would alter competitive dynamics between AI developers".[240]

Meta echoed these points. Such requirements would, it argued, create "a real risk of stifling innovation" in the UK.[241]

102. OpenAI and Google also warned of potential security implications.[242] As Google explained, "requiring the disclosure of training data as outlined in the Government's initial consultation … could equip bad actors with the knowledge to attack the model".[243]

103. Finally, stakeholders questioned whether granular disclosures would achieve their desired aim. Meta took the view that "such proposals would accomplish little practical benefit".[244] techUK told us that "lengthy and exhaustive disclosures could make it harder for smaller rightsholders to check what is being trained on and what isn't."[245] Ms Ali stated simply: "If you are not able to offer transparency in order to meet the individual, granular remuneration for rightsholders, what is the point of it?"[246] OpenAI argued that AI firms "already provide significant insights into how models are trained—often through documents like 'model cards'".[247]

104. techUK told us that its members "hold diverse perspectives on transparency measures and this is not a settled area". It emphasised the need to avoid "a one-size-fits-all approach".[248] There was consensus among the technology sector evidence to our inquiry, however, that general, high-level summaries of training data are the preferred option for AI developers. Google conceded that these "may be feasible",[249] while OpenAI described "aggregate, non-specific disclosures" as "a more balanced approach".[250] Meta noted that

237 Q 45 (Ed Newton-Rex)
238 Written evidence from Dr Andres Guadamuz (AIC0026)
239 *Ibid.*
240 Written evidence from OpenAI (AIC0004)
241 Written evidence from Meta (AIC0009)
242 Written evidence from OpenAI (AIC0004) and Meta (AIC0009)
243 Written evidence from Google (AIC0010)
244 Written evidence from Meta (AIC0009)
245 Written evidence from techUK (AIC0011)
246 Q 71 (Vinous Ali)
247 Written evidence from OpenAI (AIC0004)
248 Written evidence from techUK (AIC0011)
249 Written evidence from Google (AIC0010)
250 Written evidence from OpenAI (AIC0004)

these "could also include information regarding web crawlers used to collect data for AI training".[251]

105. Commenting on the EU's approach of requiring public, "sufficiently detailed" summaries of model training data, CREATe said that "the transparency mechanism envisaged by the EU AI Office template is promising", but that "the level of granularity required needs to be settled in negotiations with industry."[252] Mr Sinclair told us:

> "The EU's approach of a summary could get you to a good answer, but the EU's implementation of that is really struggling. UK policymakers should be very careful about thinking that the EU is further along than it is, as it has not managed to land this yet."[253]

Both Meta and OpenAI cautioned against adopting the EU's approach, warning that this could "reduce competitiveness without creating material benefit",[254] and "lead to a significant setback in achieving the Government's AI-driven growth goals".[255]

106. While our evidence focused on the public reporting element of the EU's approach, we note that the EU General-Purpose AI Code of Practice also includes a non-public "model documentation form" through which developers can disclose more granular information about training data and methods to downstream model providers, the AI Office or national competent authorities, subject to confidentiality safeguards.[256] This type of confidential disclosure mechanism could, in our view, offer a compromise, by providing a route for rightsholders to query via a regulator whether their content has been used for AI training, without the need for public disclosure of the whole training dataset. The Secretary of State for Culture, Media and Sport and the Rt Hon Liz Kendall MP, Secretary of State for Science, Innovation and Technology, suggested that they were willing to give this idea further consideration.[257]

Voluntary vs mandatory transparency requirements

107. The Government's AI and copyright progress update reported strong support for statutory transparency measures among creative sector respondents to its consultation.[258] The Secretary of State for Culture, Media and Sport told us that the recent technical working groups "have also surfaced the need for transparent processes, underpinned by legislation."[259]

108. This was echoed in our evidence, though witnesses were also keen to emphasise that this would not necessitate reform of existing copyright law.[260] The Creative Rights in AI Coalition stated that requirements "are unlikely

---

251 Written evidence from Meta (AIC0009). See paras 143–58 for further discussion of web crawlers.
252 Written evidence from CREATe Centre, University of Glasgow (AIC0007)
253 Q 75 (Matthew Sinclair)
254 Written evidence from Meta (AIC0009)
255 Written evidence from OpenAI (AIC0004)
256 See Box 3.
257 Q 150 (Secretary of State for Science, Innovation and Technology, Secretary of State for Culture, Media and Sport)
258 HM Government, *Copyright and artificial intelligence statement of progress under Section 137 Data (Use and Access) Act*, 15 December 2025, p 11
259 Q 139 (Secretary of State for Culture, Media and Sport)
260 Written evidence from Creative Rights in AI Coalition (AIC0001) and Alliance for Intellectual Property (AIC0002)

to be effective without primary legislation, as there are strong incentives for AI firms not to participate in a voluntary model."[261] The British Copyright Council argued that "previous attempts at voluntary measures have largely failed".[262] Commenting on the voluntary code of practice introduced in the EU, Professor Eleonora Rosati, Professor of Intellectual Property Law at Stockholm University, stated:

> "In my view, a voluntary obligation is no obligation. If the EU legislator—in this case—has decided to impose a transparency requirement, that requirement should be complied with … otherwise there is a risk of watering down any meaning and, indeed, the concept of an obligation on transparency."[263]

109. The Copyright Licensing Agency suggested that only clear statutory duties on AI developers would create the "fair playing field" needed for a viable UK licensing market to emerge.[264] Mr Newton-Rex observed that, currently, companies that look to license their training data can face a competitive disadvantage:

> "I know specific instances where fairer companies have gone to an investor, who has said, 'We cannot invest in you because you license your training data. How are you going to compete with the companies that don't? You're spending more money'."[265]

110. The British Copyright Council told us that any new legislation should reference that there would be "clear consequences for AI developers ignoring transparency obligations", which might include fines or market restrictions. It highlighted the need for "a regulatory body which will support both a renewed focus on compliance with, and practical enforcement of, existing copyright law provisions."[266]

111. Our evidence did not identify which body would be best placed to monitor compliance with any future transparency requirements. We note, however, the Secretary of State for Science, Innovation and Technology's observation that Ofcom may not have the resource to take on such responsibilities.[267] We note too that Lord Patrick Vallance, then Government Chief Scientific Adviser, identified in his 2023 review a role for the Intellectual Property Office (IPO) in this area:

> "To increase confidence and accessibility of protection to copyright holders of their content as permitted by law, we recommend that the Government requires the IPO to provide clearer guidance to AI firms as to their legal responsibilities, to coordinate intelligence on systematic copyright infringement by AI, and to encourage development of AI tools to help enforce IP rights."[268]

---

261 Written evidence from Creative Rights in AI Coalition (AIC0001)
262 Written evidence from British Copyright Council (AIC0006)
263 Q 29 (Prof Eleonora Rosati)
264 Written evidence from Copyright Licensing Agency (AIC0005)
265 Q 36 (Ed Newton-Rex)
266 Written evidence from British Copyright Council (AIC0006)
267 Q 150 (Secretary of State for Science, Innovation and Technology)
268 HM Government, *Pro-innovation Regulation of Technologies Review—Digital Technologies*, March 2023, p 9

112. According to the Government's recent progress statement, consultation responses from technology sector respondents were more divided. Many preferred "non-legislative approaches" or "light-touch regulation". The update highlighted that respondents across sectors stressed that any future regime should avoid unnecessary administrative and financial burdens.[269] In a similar vein, CREATe cautioned that "as a public law intervention, the regulation of transparency comes with bureaucracy, as well as compliance and enforcement costs." As such, it recommended "exploring transparency obligations that are privately enforceable."[270]

113. Dr Alina Trapova, Lecturer in Law at University College London, noted that she had been "very sceptical" of the "voluntary nature" of the code of practice introduced in the EU. She observed, however, that "it appears that the [AI] companies are adhering to it". She explained that "once you, as an AI company, sign the code of practice, it is presumed that you have complied with Article 53 of the AI Act and the copyright-related obligations." As such, in her view, this approach could be "a good way of getting compliance with these complicated provisions".[271]

114. **Meaningful transparency reporting on AI training data and data use practices is a prerequisite for rebuilding trust between the AI and creative sectors and for enabling effective licensing and enforcement. Voluntary codes alone are unlikely to drive the level of compliance required to address rightsholders' concerns and create a level playing field for developers.**

115. **Our evidence was clear that high-level, aggregate transparency disclosures are unlikely to meet rightsholders' needs and that more granular transparency reporting is required. However, in their evidence to us, AI developers objected to the prospect of granular transparency reporting, highlighting technical barriers, commercial sensitivity, competitive dynamics and compliance costs. We are persuaded that reaching a solution that meets the needs of both sides is feasible. The Government has a clear role to play in facilitating discussions between AI firms and rightsholder representatives to develop proportionate, workable proposals.**

116. ***The Government should seek to set best practice, working in collaboration with international partners, by developing a clear, proportionate framework for transparency reporting on data use in AI training. This should go beyond existing voluntary practice and improve on the high-level summaries required under current EU provisions, while remaining mindful of the potential impact on the UK's AI ecosystem.***

117. ***Transparency obligations for large AI developers should be given statutory weight. The Government should identify the appropriate regulatory body to set standardised reporting formats, monitor compliance and take enforcement action as required. In designing such obligations, particular care should be taken to avoid***

269 HM Government, *Copyright and artificial intelligence statement of progress under Section 137 Data (Use and Access) Act*, 15 December 2025, p 11

270 Written evidence from CREATe Centre, University of Glasgow (AIC0007)

271 Q 29 (Dr Alina Trapova)

disproportionate burdens on smaller UK-based firms, for example through tailored requirements.

118. ***The Government should also explore the potential for requiring commercial AI developers to make more granular confidential disclosures about their training data and methods to the relevant regulator, along the lines of the model documentation form incorporated into the EU's General-Purpose AI Code of Practice, so as to balance developers' commercial interests with rightsholders' needs for meaningful transparency.***

Territorial scope

As noted in Chapter 2, large-scale generative AI model training does not typically take place in the UK.[272] Mr Sinclair observed that this reality limits the practical impact of any new domestic transparency regime: even if the UK were to introduce new requirements, "no one is training here anyway, so no one is filling out our transparency forms".[273] The Secretary of State for Science, Innovation and Technology similarly noted that, while the Government is "open" to proposals to improve transparency for rightsholders, many larger companies may regard such obligations as unduly burdensome and simply continue to develop and train models in other jurisdictions, "where we do not have control".[274] As Antony Walker, Deputy Chief Executive Officer of techUK, put it:

> "If I have to give away all the fundamental IP about how my model is built if I train and license in the UK, but I do not have to do that in the US, what would my legal counsel say? … 'Go and train the model in the US'."[275]

119. In response, rightsholder organisations argued that transparency obligations should apply to the placing of AI systems on the UK market, rather than only to the location of model training.[276] The BCC warned that, without transparency obligations that are effective in the UK, "the Government's ambition of creating growth through AI development in the UK cannot be fulfilled as AI companies will sidestep the obligations by residing abroad."[277]

120. Mr Meredith likened this to existing regulatory approaches for physical goods.[278] He argued that "any trader wishing to sell physical goods in the UK market … must comply with UK standards. There is no reason AI developers should be treated differently, or why the Government should weaken standards resulting in UK consumers accessing inferior products and services."[279]

121. The EU has sought to give effect to such provisions in the AI Act. Recital 106 states that the copyright-related obligations in Article 53 apply even where the relevant TDM activities occur outside the EU. The stated aim is to ensure a "level playing field" and prevent non-EU providers gaining

272 See para 18.
273 Q 70 (Matthew Sinclair)
274 Q 150 (Secretary of State for Science, Innovation and Technology)
275 Q 72 (Antony Walker)
276 Written evidence from British Copyright Council (AIC0006) and Copyright Licensing Agency (AIC0005)
277 Written evidence from British Copyright Council (AIC0006)
278 Q 8 (Owen Meredith) and supplementary written evidence from News Media Association (AIC0022)
279 Supplementary written evidence from News Media Association (AIC0022)

a competitive advantage by training models under more lenient copyright standards in other jurisdictions.[280] In external commentary, Dr João Pedro Quintais has described this as a "product safety" approach:[281] any general purpose model made available in the EU is expected to be "copyright-safe",[282] and providers must have internal policies to respect EU copyright rules even where training occurred abroad.

122. Academic commentators have questioned whether this approach will in fact achieve its intended result. Dr Trapova stressed that copyright is 'territorial': each country sets and enforces its own copyright rules, and those rules do not usually govern acts that take place entirely abroad.[283] As Dr Adam Buick notes, if all of the copying and TDM needed to train a model is carried out in a country that allows such uses without permission, then, under ordinary copyright principles, there is no infringement in that country or within the EU, even if the trained model is subsequently offered to EU users. He argues that this creates a paradoxical incentive: a provider can comply with EU copyright law, and avoid EU opt-out and transparency obligations, by ensuring that training data are not collected from EU-hosted servers and that all training occurs elsewhere, potentially leaving EU-based developers subject to stricter obligations than competitors in more permissive jurisdictions.[284]

123. In light of these tensions, Dr Trapova questioned whether copyright law is the appropriate vehicle for enforcing extraterritorial principles. She suggested that, where a national legislator wishes to prevent AI models trained on infringing content abroad from being deployed domestically, an alternative could be to rely on unfair competition law. In this approach, an AI company operating in the UK with a model that has been "unfairly" trained (by using unlicensed content in ways that would infringe if done in the UK) could be treated as having obtained an unfair competitive advantage over firms that have cleared licences or trained "fairly". She cautioned, however, that this would be "ambitious" in the UK context, which has historically resisted the introduction of a general unfair competition regime and has relied instead on doctrines such as passing off.[285] Mr Meredith suggested that the Government could address this issue through non-legislative means, for example by making compliance with UK copyright law a "key factor" in its procurement decisions.[286]

124. We heard that ambitious, far-reaching transparency regimes could affect not only how and where AI models are developed, but whether they are made available in the UK at all. Meta stated:

---

280 European Commission, *AI Act Explorer: Recital 106*

281 João Pedro Quintais, 'Copyright, the AI Act and Extraterritoriality', *Policy Brief—The Lisbon Council* (June 2025): https://dx.doi.org/10.2139/ssrn.5316132

282 Supplementary written evidence from Dr Alina Trapova (AIC0021)

283 *Ibid.*

284 Dr Adam Buick, 'Copyright and AI training data—transparency to the rescue?', *Journal of Intellectual Property Law & Practice*, vol 20, Issue 3 (March 2025), pp 182–92: https://doi.org/10.1093/jiplp/jpae102

285 Article 53(1)(c) "effectively assumes that developers training models outside the EU will nonetheless choose to align their practices with EU copyright standards, rather than risk exclusion from such a significant market. The assumption, however, does not always hold. Developers may conclude that compliance costs or regulatory uncertainty outweigh the benefits of EU market access, leading them to withhold certain products altogether". Noam Shemtov, 'Against Extraterritorial Reach: Why the UK Should Resist Extending AI Regulation to Overseas Model Training', *Queen Mary Law Research Paper No. 482/2026* (January 2026): https://dx.doi.org/10.2139/ssrn.6102107

286 Supplementary written evidence from News Media Association (AIC0022)

> "Jurisdictions that impose burdensome, granular transparency requirements that do not sufficiently protect business confidential information and trade secrets should expect those requirements to disincentivise developers building and releasing their models."[287]

125. When considering this risk, witnesses expressed divergent views. Mr Newton-Rex told us that "if an AI company has stolen half the world's work, I am perfectly fine with putting them off coming here".[288] By contrast, Guy Gadney, CEO of the createch business Charismatic.ai, noted that companies based overseas often enjoy access to leading frontier models weeks before they are made available in the UK, adding: "in this world two weeks is a long time to be able to get a competitive foothold". He cautioned against legislation that may inadvertently prejudice either the nascent UK creative AI sector or the wider creative industries by limiting their access to frontier technologies.[289]

126. **We heard compelling evidence that any UK transparency regime for AI should, in principle, apply to all models available on the UK market, regardless of where they are trained. At the same time, it is important to recognise the territorial nature of copyright and the limits of its ability to regulate training that occurs entirely overseas.**

127. ***In giving effect to that principle, the Government should consider how public procurement and regulatory tools could support compliance with UK transparency requirements by AI developers operating in the UK. At a minimum, it should design transparency requirements in a way that minimises incentives for UK-based developers to relocate training abroad, or for frontier model providers to delay or withhold the release of new models in the UK, which could be detrimental to UK creators, innovators and consumers.***

Sovereign AI

128. Witnesses also suggested that the Government could pursue a strategy to build responsibly-trained domestic AI capability and reduce reliance on imported models by focusing on 'sovereign AI'.[290] The Government defines sovereign AI as the UK's ability "to access, influence or control the development and deployment of critical capabilities to protect our national interests and unlock economic growth".[291]

129. The Government has committed significant resources in pursuit of AI sovereignty, including the establishment of the Sovereign AI Unit. This will aim to "build and harness the UK's AI capabilities"[292] through a strategy involving £2 billion in investments, ambitious plans for compute infrastructure, proposals for new AI Growth Zones, and international collaborations such as the US-UK Technology Prosperity Partnership.[293] The Secretary of State for Science, Innovation and Technology told us:

---

287 Written evidence from Meta (AIC0009)
288 Q 48 (Ed Newton-Rex)
289 Q 124 (Guy Gadney)
290 Q 124 (Guy Gadney), written evidence from Creative Rights in AI Coalition (AIC0001) and written evidence from Copyright Licensing Agency (AIC0005)
291 HC Written Answer, 59030, Session 2024–26
292 HM Government, *Sovereign AI Unit*, 17 July 2025
293 ai@cam, *Navigating AI sovereignty: strategic choices for the UK*, November 2025

> "I really do want to see great British companies and great British ideas be backed … It is healthy for this country and for the world to have a strong, proud British tech sector, based on our values, to help lead the world and provide greater competition."[294]

130. The Creative Rights in AI Coalition argued that a sovereign UK AI industry should be built in partnership with UK creators and rightsholders to drive growth for businesses, consumers and citizens. In its view, this objective will not be achieved by offering "unchecked market access or free data" to predominantly US-owned firms, but by enabling the UK to play to its strengths through closer engagement with domestic AI companies and UK businesses.[295] Other witnesses identified the UK's "gold-standard" copyright framework and high-quality, creative data as key national strengths that should form part of any sovereign AI strategy.[296]

131. Other countries are already experimenting with more transparent, domestically governed approaches to AI development. The Swiss AI Initiative's Apertus LLM offers a high level of transparency about its architecture, training process and data, with model weights, intermediate checkpoints, documentation and details of the training datasets made publicly available.[297] It has been described by its developers as a "blueprint" for how a trustworthy, sovereign and inclusive model can be developed while adhering to European copyright rules.[298] We were also encouraged by the Seattle-based Allen Institute for AI's OLMo project. The model weights, training code, data pipelines and intermediate checkpoints for its family of open LLMs are released alongside an evaluation suite, with the stated aim of enabling independent scrutiny and reuse of the entire "model flow" rather than just the final parameters.[299]

132. While it should be noted that the training data for both of these projects is likely to include a large number of copyrighted works, their greater transparency should enhance rightsholders' ability to seek remuneration for the use of their content.[300] Guy Gadney, Chief Executive Officer of Charismatic.ai, argued that the UK should similarly support the development of open or sovereign models trained on properly sourced and remunerated material, as a matter of policy choice. He felt this would allow the UK to support innovation in priority sectors such as the creative industries, even if these

---

294 Q 151 (Secretary of State for Science, Innovation and Technology)

295 Written evidence from Creative Rights in AI Coalition (AIC0001)

296 Q 125 (Guy Gadney) and written evidence from Copyright Licensing Agency (AIC0005)

297 The model architecture, training code, model weights and intermediate checkpoints are all released under a permissive open-source licence; the training dataset is built only from publicly available data and is filtered to respect machine-readable opt-out requests from websites, even "retroactively", with further filtering to remove personal data and other undesired content before training. Project documentation explicitly states that development complied with copyright law and with transparency obligations under the EU AI Act. Project Apertus, 'Apertus: Democratizing Open and Compliant LLMs for Global Language Environments', (2025): https://doi.org/10.48550/arXiv.2509.14233

298 AI News, *Switzerland releases 100% open AI model*, September 2025

299 The Allen Institute for Artificial Intelligence, *Olmo 3: Charting a path through the model flow to lead open-source AI*, 20 November 2025

300 The Apertus model was trained using a series of web-crawl datasets, using an "opt-out policy" based on robots.txt: Project Apertus, 'Apertus: Democratizing Open and Compliant LLMs for Global Language Environments', (2025): https://doi.org/10.48550/arXiv.2509.14233. The OLMo model was trained on Dolma, a 3-trillion-token dataset, which includes significant portions of web-crawled data: The Allen Institute for Artificial Intelligence, *Ai2 Dolma: 3 trillion token open corpus for language pretraining*, 18 August 2023

models do not immediately match the general capabilities of the largest commercial systems.[301]

133. Owen Meredith, Chief Executive Officer of the News Media Association, suggested that "the UK should have confidence … in the strength" of both its AI sector and the creative industries. He added:

> "The UK can have its own legal system and exercise its own jurisdiction in a way that supports both sectors to grow alongside each other … There are plenty of examples where we have demonstrated a UK-first agenda and approach, and I think that we can continue to do that in this space."[302]

134. **International examples such as the Swiss AI Initiative's Apertus system and the Allen Institute for AI's OLMo project indicate that it is possible to build competitive models whose training data and development pipelines are open to external scrutiny. Any UK approach to sovereign AI should draw on these lessons and align future capability with robust standards of compliance, accountability and respect for copyright.**

135. ***The UK should not resign itself to long-term dependence on opaque models trained overseas. Instead, the Government's sovereign AI efforts should prioritise the development and adoption of domestically governed models with transparency built in by design, including clear information about training data and development processes. The Government must be clear that where the use of copyrighted data in the training of such models is not covered by an existing exception, remuneration for rightsholders will be provided.***

301 Q 124 (Guy Gadney)
302 Q 2 (Owen Meredith)

CHAPTER 4: EMERGING TECHNICAL SOLUTIONS

136. Alongside transparency to help them understand how their works have been used by AI models, rightsholders want ways to exercise control over those uses and to share in the value created by AI outputs—technical mechanisms that allow them to state clearly whether, and on what terms, their content may be used by AI systems, and to participate in the revenues that follow.

137. In this chapter, we examine existing 'rights-reservation' or 'opt-out' mechanisms. We consider how emerging and forward-looking frameworks may address their limitations over time, both in enabling more effective control and in supporting licensing and value-sharing arrangements for AI uses of creative content. We also assess the workability of introducing a new commercial text and data mining (TDM) exception with a rights-reservation mechanism, as initially proposed by the Government's consultation, against this backdrop. Finally, we look at methods to support clear labelling of AI-generated material so that audiences can distinguish it from human-authored work.

Challenges for rights reservation

138. Assessing the effectiveness of existing rights-reservation mechanisms requires an understanding of the obstacles rightsholders may face when trying to assert and enforce their rights. For this purpose, it is helpful to treat creative works as digital 'assets': individual items of content, such as images, videos, audio recordings or text files, that can be copied, shared and processed independently across digital environments.[303]

139. A research report by CoSTAR National Lab and partners, *Time to ACCCT: an AI copyright framework for UK creative industries*, highlights two related sources of complexity. First, a single asset may combine multiple media types, each subject to different rights and ownership arrangements. For example, an online news article may incorporate images or video produced and owned by others. Second, many assets carry multiple underlying intellectual property rights, administered through layered arrangements. In music, for instance, a single pop song typically involves contributions from producers, songwriters, composers, artists and session musicians, with rights managed across labels, publishers and collective management organisations. The ACCCT report authors conclude that in this context intellectual property:

> "should be viewed as organised into 'bundles of rights' that require a high degree of granularity. A framework that seeks to promote transparency and consent in this context would need to be able to identify, track and attribute across these bundles and their use in different creative sectors."[304]

140. Oliver Ilott, Interim Director General for Artificial Intelligence at the Department for Science, Innovation and Technology, pointed out, however, that digital material may not incorporate rights or licensing information at this level of specificity. He gave the example of a book review blog post, which might include quotes from a novel, as well as text by the blogger, but without this varying copyright ownership being reflected in the site's

303 C2PA Specifications, *C2PA and Content Credentials Explainer*

304 CoSTAR National Lab, DECaDE and Sheridans, *Time to ACCCT: an AI copyright framework for UK creative industries*, 21 May 2025, p 16, hereafter referred to as CoSTAR National Lab, *Time to ACCCT*

metadata. As he explained, this can make it difficult to track rightsholders preferences about whether their work may be accessed for AI training: "If that author has exercised an opt-out somewhere, how do you know that it attaches to the blog that I might then have published?" [305]

141. We also heard that creative assets rarely remain confined to their original context. Tom Kiehl, Chief Executive Officer of UK Music, told us that once a track is uploaded to one platform it may then appear on multiple user-generated content services and in many different formats. He explained that these "downstream copies" are "very challenging" for rightsholders to track or control, particularly in relation to scraping for AI training.[306]

Existing rights-reservation tools

142. Technologies for expressing creators' preferences on whether their content is used by AI systems can broadly operate at two levels: site-based or location-level mechanisms that apply rules to all assets under a web domain or path, and unit- or asset-level mechanisms that attach permissions and restrictions directly to individual copyrighted works.[307] Each approach has distinct strengths and limitations in terms of scope and persistence through the AI lifecycle.

Site-based or location-level controls

143. Most of the technical tools currently available to rightsholders are site-based controls. These allow a creator or content host to apply a blanket permission or restriction to assets located on all or part of a website.[308] The main mechanism is the Robots Exclusion Protocol (REP), implemented through a file named 'robots.txt' placed at the site root. In simple terms, this file gives automated bots a yes/no instruction about whether they may access particular parts of a site. These instructions have long been used to control how search engines crawl and index web pages.[309] They are also read by bots when content is crawled and scraped for use in AI training or retrieval systems.[310]

144. To support the TDM opt-outs introduced in the EU,[311] the World Wide Web Consortium (W3C) has developed a complementary standard, the Text and Data Mining Reservation Protocol (TDMRep). TDMRep provides a machine-readable way for web publishers to reserve their TDM rights and to link to licensing policies using a "tdm-reservation" signal that can be applied at site or path level.[312] We discuss the EU's approach to rights reservation later in this chapter.

---

305 Q 140 (Oliver Ilott)
306 Q 2 (Tom Kiehl)
307 Q 89 (Prof John Collomosse) and written evidence from Microsoft (AIC0012). See also Kar Balan, Andrew Gilbert, John Collomosse, 'Content ARCs: Decentralized Content Rights in the Age of Generative AI', *International Conference on AI and the Digital Economy (CADE '25)*, (2025): https://doi.org/10.48550/arXiv.2503.14519
308 Kar Balan, Andrew Gilbert, John Collomosse, 'Content ARCs: Decentralized Content Rights in the Age of Generative AI', *International Conference on AI and the Digital Economy (CADE '25)*, (2025): https://doi.org/10.48550/arXiv.2503.14519
309 Written evidence from Microsoft (AIC0012)
310 Michael Dinzinger, Florian Heß and Michael Granitzer, 'A Survey of Web Content Control for Generative AI' (2024): https://arxiv.org/html/2404.02309v1
311 See Table 2.
312 Michael Dinzinger, Florian Heß and Michael Granitzer, 'A Survey of Web Content Control for Generative AI' (2024): https://arxiv.org/html/2404.02309v1

145. OpenAI, Google and Meta argued that robots.txt and TDMRep provide effective ways for rightsholders to express their preferences about how their works are accessed by web crawlers. They characterised these mechanisms as simple, scalable, universally understood and widely adopted.[313] Microsoft similarly recognised that many publishers currently rely on robots.txt as a location-based way of signalling their preferences. It acknowledged, however, that "because robots.txt was designed for web crawling rather than AI preference", work is underway to develop mechanisms that would enable content owners to express more nuanced rights positions.[314]

146. The REP does not differentiate between the different purposes for which data may be accessed.[315] In response, some web crawl operators have separated their activities across multiple bots and identified each one with a distinct 'user-agent' in robots.txt. Major search and AI providers typically operate multiple crawlers linked to particular services or products.[316] In principle, this enables publishers to allow general search indexing while restricting other uses, such as AI training or 'AI answer' services, within the same robots.txt framework.

147. We heard, however, that when crawlers are not separated, robots.txt "forces a binary choice: allow access for indexing and discovery, or block access entirely." For many publishers, reserving rights against AI training or AI-generated summaries integrated into search results can therefore mean sacrificing visibility in traditional search.[317]

148. Witnesses from both creative and technical backgrounds highlighted further limitations of these site-level tools. First, as they are inherently location-based, they can be a poor fit for complex rights structures, such as web pages that combine multiple assets with different owners.[318] Professor John Collomosse, Professor of Computer Vision and AI at the University of Surrey, told us that since the preferences attached to a work uploaded to a platform are "not going to move with it" when that work is "reposted or reshared elsewhere on the internet", this also limits the usefulness of site-level tools for enforcing consent in the context of generative AI.[319] Ed Newton-Rex, Chief Executive Officer of Fairly Trained, likewise pointed out that many creators' works appear primarily as downstream copies on third-party platforms where the platform, not the creator, controls robots.txt, meaning that individual rightsholders cannot use these tools to reserve rights in practice.[320]

149. Witnesses also stressed that robots.txt and TDMRep operate on a voluntary basis. They rely on crawlers identifying themselves and complying with the

313 Written evidence from OpenAI (AIC0004), Meta (AIC0009) and Google (AIC0010)
314 This includes updates to the REP being discussed at the Internet Engineering Task Force (IETF) and emerging content-based controls such as the International Standard Content Code (ISCC). Written evidence from Microsoft (AIC0012)
315 Enze Liu et al, 'Somesite I Used to Crawl: Awareness, Agency and Efficacy in Protecting Content Creators From AI Crawlers' (2024): https://doi.org/10.48550/arXiv.2411.15091
316 Google, for example, documents multiple distinct crawlers. Google for Developers, *Google's common crawlers*, 11 February 2026
317 Written evidence from RSL Collective (AIC0023)
318 See para 140.
319 Q 89 (Prof John Collomosse)
320 Q 41 (Ed Newton-Rex)

rules they encounter.[321] Studies suggest that while the majority of the AI crawlers operated by large companies do respect robots.txt, some ignore these signals or even attempt to circumvent website preferences using 'stealth' user agents.[322]

150. Google and Microsoft pointed to their own product-specific controls, which build on robots.txt and TDMRep, as evidence that rightsholders already have meaningful ways to manage AI use of their content.[323] In 2023, Google introduced a specialised user agent, "Google-Extended", as a "new control that web publishers can use to manage whether their sites help improve" its generative AI services Bard and Vertex AI. Both products use retrieval-augmented generation (RAG) or 'grounding' to connect their outputs to up-to-date sources of information.[324]

151. During our future of news inquiry, Google told us that Google-Extended provides "an effective way for publishers to exclude their content from this grounding service without opting out of Google Search".[325] It stated again in our current inquiry that:

> "Any web publisher can implement Google-Extended to opt out of model training. … it does not impact a site's inclusion or ranking in Google Search."[326]

152. Despite these assurances, witnesses told us that publishers cannot realistically refuse certain AI uses without facing practical penalties in Google Search. They told us that Google currently uses a single primary crawler, Googlebot, for both general search indexing and accessing content for use in its AI search interfaces through RAG.[327] Mr Meredith told us that "Google is able to effectively leverage its dominance and force news media publishers into giving access to their content for use in their AI Overviews and AI Mode summaries", which reduces their web traffic,[328] or "risk being de-ranked in Search" and lose traffic from Google altogether.[329]

153. According to the internet services company Cloudflare, this "lack of crawler separation" means Google "is not competing on a level playing field with other

321 Q 89 (Prof John Collomosse) and written evidence from Prof Sylvester Kaczmarek (AIC0016). See also Enze Liu et al, 'Somesite I Used to Crawl: Awareness, Agency and Efficacy in Protecting Content Creators From AI Crawlers', (2024): https://doi.org/10.48550/arXiv.2411.15091; Taein Kim et al, 'Scrapers Selectively Respect robots.txt Directives: Evidence From a Large-scale Empirical Study', *IMC '25: Proceedings of the 2025 ACM Internet Measurement Conference*, (2025), pp 541–557: https://doi.org/10.1145/3730567.3764471

322 Cloudflare has alleged that the AI search company Perplexity has engaged in "stealth crawling": using undeclared useragents, rotating IP addresses and bypassing robots.txt to access sites that had attempted to block it, including test domains set up specifically to detect such behaviour. Cloudflare, *Perplexity is using stealth, undeclared crawlers to evade website no-crawl directives*, 4 August 2025. Perplexity has disputed these claims. TechRadar, *Perplexity hits back after Cloudflare slams its online scraping tools*, 6 August 2024

323 Written evidence from Google (AIC0010) and Microsoft (AIC0012). See also Microsoft Bing Blogs, *Announcing new options for webmasters to control usage of their content in Bing Chat*, 22 September 2023

324 Vertex AI is Google's machine learning platform for deploying generative AI-powered search and chat applications, and Bard, now rebranded as Google Gemini, is Google's conversational AI assistant. Google, *An update on web publisher controls*, 28 September 2023

325 Correspondence, *Director, Government Affairs and Public Policy, Google to the Chair of the Communications and Digital Committee*, 20 November 2024

326 Written evidence from Google (AIC0010)

327 Q 81 (Ed Conolly), written evidence from Cloudflare (AIC0015) and supplementary written evidence from News Media Association (AIC0022)

328 See para 45.

329 Q 13 (Owen Meredith) and supplementary written evidence from News Media Association (AIC0022)

AI companies".[330] Cloudflare's data indicated that around 8% of publishers fully block Google's crawler, compared with 78% and 80% blocking the equivalent crawlers for ChatGPT and Anthropic, reflecting publishers' dependence on Google Search traffic. It argued that tying search and AI use together in a single crawler also undermines emerging monetisation solutions, such as services that enable publishers to license their content to AI companies on negotiated terms.

154. In October 2025, the Competition and Markets Authority (CMA) designated Google with 'strategic market status' (SMS)[331] in search services under the Digital Markets, Competition and Consumers Act 2024.[332] This designation enabled the CMA to impose targeted "conduct requirements" on Google to secure fair dealing, open choices, and trust and transparency in the market for search. Following the designation, the CMA has proposed a package of requirements for Google's search services in the UK and is consulting on their implementation.[333]

155. The proposed measures include a set of "publisher controls", aimed at ensuring that:

> "Content publishers obtain a fairer deal by giving them more choice and transparency over how their content is used in Google's AI Overviews. Publishers will be able to opt out of their content being used to power AI features such as AI Overviews or to train AI models outside of Google search. Google will also be required to take practical steps to ensure publisher content is properly attributed in AI results."[334]

156. Writing to the Committee after the announcement, Roxanne Carter, Global IP Lead for the Government Affairs and Policy team at Google, confirmed that Google is:

> "now exploring updates to our controls to let sites specifically opt out of Search generative AI features. Our goal is to protect the helpfulness of Search for people who want information quickly, while also giving websites the right tools to manage their content."[335]

157. Several witnesses were strongly supportive of the CMA's decision to designate Google with SMS in search, and urged the CMA to use these powers to secure meaningful publisher controls in practice.[336] The NMA, in particular, called for remedies to ensure that crawling for General Search and crawling for AI Search Interfaces are kept separate. It also warned, however, that the behavioural remedies currently proposed by the CMA may not go

---

330 Written evidence from Cloudflare (AIC0015)

331 "To designate a firm with SMS, the CMA must decide that it has substantial and entrenched market power and a position of strategic significance in a digital activity. Google having SMS does not imply that it has acted anti-competitively". Competition and Markets Authority, *Press release*: CMA proposes package of measures to improve Google search services in UK, 28 January 2026

332 Competition and Markets Authority, *CMA confirms Google has strategic market status in search services*, 10 October 2025

333 Competition and Markets Authority, *Press release*: CMA proposes package of measures to improve Google search services in UK, 28 January 2026

334 *Ibid.*

335 Supplementary written evidence from Google (AIC0024)

336 Q 83 (Eugene Huang), written evidence from Cloudflare (AIC0015) and supplementary written evidence from News Media Association (AIC0022)

far enough, as they do not require "full physical separation of crawlers for Search and individual AI products".[337]

158. **We welcome the proposals from the Competition and Markets Authority (CMA) to introduce conduct requirements for Google's search services following its designation of Google with 'strategic market status'. For these measures to address the concerns raised in our inquiry, they must ensure publishers have meaningful technical control over whether their content is used for Google Search, for AI training, and for retrieval-augmented generation or 'grounding' features across Google's products, including through clearer signalling of distinct uses. We also note more broadly the importance of the CMA's digital markets regulatory regime and will continue to scrutinise the effectiveness of its implementation.**

Unit- or asset-level controls

159. Unit- or asset-level approaches seek to embed provenance and rights information within individual works themselves in a structured, machine-readable form. Several witnesses highlighted the Coalition for Content Provenance and Authenticity (C2PA) standard as a particularly important development in this space.[338]

160. C2PA provides a framework for embedding provenance metadata—known as 'content credentials'—directly into media files. Prof Collomosse characterised C2PA content credentials as "a way of describing where content has come from, who made it and what has been done to it".[339] The standard has been extended to allow rightsholders to express basic AI-training opt-out preferences at asset level.[340]

161. While asset-level approaches can address some of the limitations of site-level mechanisms discussed above, witnesses recognised that metadata-only approaches are not sufficient on their own.[341] OpenAI noted that while it had seen interest in C2PA and related standards, "metadata can be stripped when content is rehosted, reformatted, or shared across different platforms, making it an unreliable mechanism for large-scale rights management". It pointed to the practical difficulty of ensuring that all representations of the same work carry consistent metadata. It stressed too that standards adopted today are inherently "forward-looking": copies of works already circulating online and which may have been ingested by AI systems in the past will not automatically bear the new information.[342]

162. The ACCCT report likewise cautions that simple metadata tags can be "easily stripped by content platforms such as social media sites that do not implement provenance standards". It argued that any effective framework

337 News Media Association, *NMA Welcomes CMA Proposals On Google Search But Stronger Remedies Are Needed*, 28 January 2026

338 Q 89 (Prof John Collomosse), written evidence from Google (AIC0010), written evidence from techUK (AIC0011), written evidence from Microsoft (AIC0012), written evidence from Independent Society of Musicians (AIC0014) and written evidence from RSL Collective (AIC0023)

339 Q 89 (John Collomosse)

340 CoSTAR National Lab, *Time to ACCCT*, p 25

341 Written evidence from OpenAI (AIC0004), techUK (AIC0011), Prof Sylvester Kaczmarek (AIC0016) and RSL Collective (AIC0023)

342 Written evidence from OpenAI (AIC0004)

must therefore make provenance information "durable" as content is distributed and modified.[343]

163. Prof Collomosse explained that asset-level metadata can be reinforced by watermarking, where an invisible identifier is injected into the asset, and fingerprinting, where a distinctive digital signature is derived from the asset so that it can be recognised, even if metadata has been removed. The resulting identifier can then be linked to a registry that stores the associated metadata, so that if that metadata is stripped from a copy, the original information can still be retrieved.[344]

164. Research suggests that robust provenance for AI therefore requires a combination of three mutually reinforcing "pillars": signed metadata, fingerprinting and watermarking.[345] Giving evidence to our inquiry, the Rt Hon Lisa Nandy MP, Secretary of State for Culture, Media and Sport, told us this approach was raised in one of the Government's recent technical working groups, where participants had effectively called for a "triple-lock" of "metadata, watermarking and neural fingerprinting".[346] The Digital Entertainment and Retail Association (ERA) stressed that transparency "fundamentally depends on accurate data to enable reporting, remuneration [and] copyright protection (including rights reservations)".[347] If machine-readable provenance information can reliably travel with an asset through this combination of measures, it may also help address some of the technical challenges that AI developers highlighted around providing more granular transparency on training data, as discussed in Chapter 3.[348]

165. While the developments above show promise, we also heard important caveats about what technology can currently deliver. Microsoft, for example, said that "persistent provenance information cannot be reliably embedded or preserved through routine transformation, reuse, or compilation" of software code.[349] OpenAI likewise described text as "the most technically challenging domain for provenance", noting that plain text lacks a universal file format and that watermarking techniques can produce high rates of false positives and negatives.[350]

The limitations of a TDM 'opt-out' system

166. The current lack of robust opt-out controls was a central argument in many stakeholders' objections to the proposed "option 3" put forward in the Government's AI and copyright consultation: introducing a new commercial TDM exception accompanied by a rights-reservation mechanism and

343 CoSTAR National Lab, *Time to ACCCT*, p 24
344 Q 93 (Prof John Collomosse)
345 John Collomosse and Andy Parsons, 'To Authenticity, and Beyond! Building Safe and Fair Generative AI Upon the Three Pillars of Provenance', *IEEE Computer Graphics and Applications*, vol 4, Issue 3, (2024), pp 82–90: https://doi.org/10.1109/MCG.2024.3380168
346 Q 146 (Secretary of State for Culture, Media and Sport)
347 Written evidence from ERA (AIC0020)
348 This idea is explored in CoSTAR National Lab, *Time to ACCCT*. The technical challenges we mention here refer to those described for example by Google: "This technical instability is further compounded by the legal landscape, where the absence of a centralised copyright registry and the poor quality of rights metadata make it functionally impossible to verify ownership". Written evidence from Google (AIC0010)
349 Written evidence from Microsoft (AIC0012)
350 Written evidence from OpenAI (AIC0004)

supporting transparency measures.[351] As noted in Chapter 1,[352] in its December progress statement, the Government reported that only a tiny minority of respondents to its consultation supported option 3, or a broader commercial TDM exception with no rights-reservation at all ("option 2"). An overwhelming majority supported licensing in all cases ("option 1").[353]

167. The Government noted that this distribution partly reflected the large number of responses from individual creators and creative industry organisations, most of whom opposed option 3 and strongly supported a 'licences in all cases' approach. Most technology-sector respondents favoured either option 3 or option 2.[354]

168. This divergence of views was mirrored in the evidence to our inquiry.[355] Creative industry witnesses were clear that a commercial TDM exception with rights reservation would not give rightsholders the control they require. Several compared the proposal to the model under Article 4 of the European Union Copyright in the Digital Single Market Directive (CDSMD).[356]

169. The Creative Rights in AI Coalition told us that the EU opt-out regime, introduced several years before the emergence of modern generative AI systems, is "a rights reservation model by another name" that has "failed to support a strong licensing market".[357] The Copyright Licensing Agency (CLA) pointed to multiple issues associated with the EU opt-out mechanism, including the lack of standardised opt-out protocols, uncertainty about the extent to which developers might honour reservations, and limited visibility over whether and how works were being used in practice.[358]

170. Concerns about the EU experience were not limited to creative-sector organisations. OpenAI told us that the opt-out framework in the EU faces "significant implementation challenges", highlighting the absence of clear technical standards for expressing reservations. In its view, this has led to "considerable variance in how rightsholders attempt to reserve their rights",

---

351 IPO, *Copyright and AI: Consultation*

352 See para 10.

353 HM Government, *Copyright and artificial intelligence statement of progress under Section 137 Data (Use and Access) Act*, 15 December 2025, p 11

354 *Ibid.*, p 11

355 For witnesses who supported strengthening licensing and/or opposed the introduction of a new commercial TDM exception with an opt-out-based rights-reservation mechanism, see QQ 4 (Tom Kiehl), 16 (Owen Meredith), 41 (Ed Newton-Rex), 50 (Reema Selhi), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from Alliance for Intellectual Property (AIC0002), written evidence from Copyright Licensing Agency (AIC0005), written evidence from British Copyright Council (AIC0006), written evidence from CREATe Centre, University of Glasgow (AIC0007), written evidence from Creative Industries Policy and Evidence Centre (AIC0008), written evidence from Independent Society of Musicians (AIC0014), written evidence from Directors UK and British Equity Collecting Society (AIC0018) and supplementary written evidence from Society of Authors (AIC0003). For those who supported some form of a broad commercial TDM exception, with or without an opt-out mechanism, see QQ 58 (Vinous Ali), 69 (Matthew Sinclair), written evidence from OpenAI (AIC0004), written evidence from Meta (AIC0009), written evidence from Google (AIC0010), written evidence from techUK (AIC0011), written evidence from Microsoft (AIC0012) and written evidence from Dr Andres Guadamuz (AIC0026).

356 See Table 2. This provision provides a copyright exception for "reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining" (Article 4(1)), unless those works have been "expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online." (Article 4(3)).

357 Written evidence from Creative Rights in AI Coalition (AIC0001)

358 Written evidence from Copyright Licensing Agency (AIC0005)

meaning AI developers "may have difficulty in reliably identifying which works can be accessed and which are off-limits".[359]

171. Witnesses told us that, in light of these experiences, a similar TDM exception with a rights-reservation mechanism would not be workable in the UK. The Society of Authors said that it was "entirely unreasonable" to expect individuals to monitor the use of their works across "vast swathes of electronic formats" in order to exercise an opt-out.[360] The Creative Industries Policy and Evidence Centre (Creative PEC) highlighted that awareness of existing rights-reservation mechanisms is "extremely low", and cited analysis that even widely used tools are "little known and poorly understood".[361] The British Copyright Council similarly stressed that, while "effective tools and standards are key for controlling content use", their impact to date has been constrained by "patchy adoption and circumvention challenges".[362]

172. Above all, witnesses stressed that rightsholders would need to trust any technical mechanism for rights reservation under a TDM exception. However, Creative PEC summarised the evidence examined in this chapter in its statement that "at present, we are not aware that such a reliable system exists". It said that developing one would require "sustained collaboration between creative industry stakeholders and technology companies".[363]

173. As a result, creative-sector organisations urged the Government not to rely on opt-out-based rights-reservation tools as the primary means of protecting their works. Instead, they called for a clear public statement that the UK would not proceed with this proposal.[364]

174. We were reassured that the Government has started to reflect these concerns in its own position. The Secretary of State for Culture, Media and Sport told us that "there are challenges with the opt-out process that we had not anticipated or fully understood before we did the consultation and which we do not currently know how to surmount". She confirmed that the Government "does not have a workable opt-out proposal on the table".[365]

175. **A broad TDM exception with an opt-out rights-reservation mechanism, as envisaged under "option 3" of the Government's consultation, is not a viable foundation for the UK's AI and copyright regime. The tools currently available for rights reservation are fragmented, poorly understood and place an unreasonable burden on individual creators.**

176. **Experience in the EU suggests that comparable provisions have neither delivered reliable control for rightsholders nor a robust licensing market. We welcome the Secretary of State's acknowledgement that there is currently no workable opt-out proposal on the table and consider that the Government should now draw a clear line under this approach. The UK's starting point should remain a licensing-first**

359 Written evidence from OpenAI (AIC0004)
360 Supplementary written evidence from Society of Authors (AIC0003)
361 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)
362 Written evidence from British Copyright Council (AIC0006)
363 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)
364 QQ 41, 50 (Ed Newton-Rex), 50, 55 (Reema Selhi), written evidence from Creative Rights in AI Coalition (AIC0001) and supplementary written evidence from News Media Association (AIC0022)
365 Q 141 (Secretary of State for Culture, Media and Sport)

model in which permission is required to use copyright-protected works for AI training.

177. ***In the light of this, we urge the Government to follow the example of the Australian Government and make a clear public statement that it will not bring forward proposals for a new commercial TDM exception with an opt-out-based rights-reservation mechanism.***

A technical foundation for a licensing-first approach

178. Although existing rights-reservation tools are not a viable foundation for a broad TDM exception, some witnesses were clear that technical measures will nevertheless be essential in supporting attribution, transparency, licensing and remuneration in a licensing-first system. Prof Collomosse told us that, at each stage of the AI lifecycle, the challenge facing rightsholders is: "how can they meaningfully control whether their work is used and under what terms, and how can they gain an opportunity to share in the value created by the model outputs?"[366] Eugene Huang, Senior Strategy Advisor and Co-founder of ProRata.ai, likewise warned that, if creators are "just being paid on the input side", they are "not taking advantage of any of the great value that is being created by AI today". In his view, this underscores the need for mechanisms that can support remuneration linked to both inputs and outputs.[367]

179. Prof Collomosse suggested that the "best way to view the AI copyright problem is as a supply-chain problem". He suggested that open provenance standards such as C2PA, when combined with watermarking and fingerprinting, are a way of "increasing transparency in the content supply chain" by attaching durable, machine-readable signals to individual assets as they move through it. ERA told us that copyright ownership and provenance information must be created "at source" by creators and rightsholders to enable stakeholders along sector-specific supply chains to understand this data and manage it effectively.[368] Prof Collomosse explained that once provenance is established in this way, it becomes easier to confirm authenticity and "assign ownership". Rightsholders can then "set up their own preferences and licensing schemes on top of that using other standards", enabling remuneration models to be layered onto individual assets rather than relying solely on blanket licensing.[369]

180. Technical controls that incorporate licensing terms are already being operationalised. For example, the Really Simple Licensing (RSL) 1.0 standard is an open, globally interoperable protocol that allows publishers and rightsholders to state the legal terms under which automated systems such as crawlers may access or license their digital assets.[370]

181. RSL builds on existing web infrastructure such as robots.txt by allowing licensing terms to be surfaced automatically to crawlers at the point they try to access content.[371] These terms can apply either across an entire site or

366 Q 80 (Prof John Collomosse)
367 Q 80 (Eugene Huang)
368 Written evidence from ERA (AIC0020)
369 Q 89 (Prof John Collomosse)
370 Written evidence from RSL Collective (AIC0023). See also RSL, *Really Simple Licensing (RSL) 1.0 Specification*.
371 RSL, RSL, *Really Simple Licensing (RSL) 1.0 Specification: 3.1 Overview and Structure*, 10 December 2025, 10 December 2025

section of a site, or to individual assets,[372] enabling both site-wide and asset-level licensing terms to be expressed in a machine-readable way.

182. Cloudflare has now developed AI-specific bot-management tools, including a feature that identifies and actively blocks AI crawlers.[373] It is also piloting a "pay per crawl" scheme allowing publishers to charge AI companies on a per-request basis when their bots access content.[374]

183. We also heard about tools focused on attribution and value allocation that operate further down the AI value chain. Vermillio's TraceID system, for example, embeds unique identifiers into digital assets to enable real-time attribution and tracking of those assets in AI-generated outputs. Vermillio has also demonstrated that, by creating "neural fingerprints" for particular creative works or franchises, such as James Bond, it can estimate the extent to which AI-generated images or videos draw on those works.[375]

184. ProRata has developed Gist.ai, an AI-powered search engine built on a corpus of licensed publisher content. Its systems measure how much particular sources contribute to a given answer, so that revenues generated by the service can be allocated proportionately between participating publishers.[376] Mr Huang acknowledged that the company is "at the early stages of this" and working at a different scale from the largest companies. He argued, however, that it has demonstrated that "there is technology that works today" to enable attribution of value and remuneration from AI outputs.[377]

185. The ACCCT report notes that many creators want to know whether their works have already been used in model training without their knowledge or consent, and to protect their work or receive value if so.[378] This was reflected in our evidence.[379] The report acknowledges, however, that emerging solutions along the value chain are largely focused on resolving transparency issues going forward. Further work is needed on retrospective rights, it argues, but this should not prevent progress.[380] Our conclusions and recommendations in Chapter 3 about the need for statutory transparency duties therefore remain an essential complement to these emerging technical and voluntary measures.[381]

186. **Effective, machine-readable rights-reservation mechanisms will be an essential component of any sustainable licensing regime. Such tools are needed to give rightsholders granular control over how their works are used by AI systems, and to enable developers to assemble the large, diverse datasets they require while operating within the law. Existing site-level rights-reservation tools were not designed with modern AI use cases in mind and have significant limitations in this context.**

---

372 RSL, *Really Simple Licensing (RSL) 1.0 Specification: 4.8 Media and data file association*, 10 December 2025
373 Cloudflare, *Block AI Bots*, 26 September 2025
374 Cloudflare, *Introducing pay per crawl: Enabling content owners to charge AI crawlers for access*, 1 July 2025
375 The Guardian, *The platform exposing exactly how much copyrighted art is used by AI tools*, 18 October 2025
376 Professional Publishers Association, *ProRata AI secures 500+ publisher deals, powering one of the largest licensed GenAI search libraries*, 19 June 2025
377 Q 83 (Eugene Huang)
378 CoSTAR National Lab, *Time to ACCCT*, p 40
379 QQ 2 (Anna Ganley), 8 (Isabelle Doran)
380 CoSTAR National Lab, *Time to ACCCT*, p 40
381 See paras 114–17.

187. **It is encouraging to see industry bringing forward new solutions that may help to address these shortcomings. Robust provenance technologies combining signed metadata, fingerprinting and watermarking seem particularly promising. By making consent and licensing signals durable as content moves through digital supply chains, these tools could help ensure rightsholders' preferences are respected, as well as creating new opportunities for remuneration at the point of data access.**

Labelling of AI-generated outputs

188. Alongside machine-readable rights-reservation and provenance tools, witnesses stressed that any effective framework must also give people visible, user-facing information about whether content has been made by generative AI or a human being.[382] Anna Ganley, Chief Executive Officer of the Society of Authors, described how AI-generated "copycat" books can currently be listed on online retailer platforms without any indication that they are AI-generated, leaving readers "duped and disappointed" when they discover they have not bought original, human-authored work.[383] The Society of Authors told us that mandatory labelling and provenance standards for AI-generated content are therefore essential "to safeguard the integrity of human authorship and maintain consumer trust", a view shared by several witnesses.[384]

189. Ofcom's report *Deepfake Defences 2: The Attribution Toolkit* describes a "label" as a "visible and recognisable icon that signifies a given characteristic or a risk to users", designed to be "easily accessible" and to "help people understand and navigate online content."[385] Established labelling practices are now being adapted for AI-generated content, with both platforms and upstream model developers beginning to implement AI-specific labels.[386]

190. Ofcom's research found that users see particular value in AI labels where synthetic content, including deepfakes, poses a greater risk of harm, such as in relation to politics, and that labels can encourage reporting of harmful content.[387] Its report on deepfake attribution tools further highlights academic evidence that "first impressions" are hard to dislodge once formed, suggesting that clear, early labelling can help people to engage more critically with content from the outset.[388]

191. Industry witnesses pointed to emerging tools designed to support these aims. Google told us that being able to determine whether content is AI-generated is key to "helping prevent the spread of misinformation", and said it is

382 QQ 11 (Tom Kiehl), 12 (Anna Ganley), written evidence from Alliance for Intellectual Property (AIC0002), written evidence from British Copyright Council (AIC0006), written evidence from Google (AIC0010), written evidence from Independent Society of Musicians (AIC0014) and supplementary written evidence from Society of Authors (AIC0003)

383 Q 12 (Anna Ganley)

384 Q 11 (Tom Kiehl), written evidence from Alliance for Intellectual Property (AIC0002), written evidence from British Copyright Council (AIC0006), written evidence from Independent Society of Musicians (AIC0014) and supplementary written evidence from Society of Authors (AIC0003)

385 Ofcom, *Deepfake Defences 2: The Attribution Toolkit*, p 23

386 Written evidence from OpenAI (AIC0004), Meta (AIC0009), Google (AIC0010) and Microsoft (AIC0012)

387 Ofcom, *Deepfake Defences 2: The Attribution Toolkit*, p 25

388 Briony Swire-Thompson, Ullrich K H Ecker, 'Misinformation and its Correction: Cognitive Mechanisms and Recommendations for Mass Communication', *Misinformation and Mass Audiences*, December 2018: https://doi.org/10.7560/314555–013

"investing heavily" in provenance tools to achieve this.[389] It cited its SynthID toolkit, which embeds watermarks directly into AI-generated images, audio, text and video. As discussed above, such asset-level provenance approaches can also provide the technical basis for visible labelling by attaching durable machine-readable signals that indicate how content has been made. Microsoft and OpenAI described similar initiatives, and both highlighted their participation in industry bodies such as C2PA.[390]

192. Ofcom has, however, warned that AI labels cannot be treated as a proxy for reliability. Its research found that 59% of users said that they may not trust content carrying an AI label, and that labels can create an "implied truth effect" in which unlabelled content is assumed to be authentic even when it may be deceptive.[391] Google similarly stressed that "not all AI-generated content is deceptive and not all deceptive content is AI-generated".[392] techUK argued that strengthening media literacy and enforcing existing legal protections will be essential to complement technological solutions.[393]

193. Ofcom likewise recognised that it is not reasonable to place the burden of detecting misleading content, particularly AI-generated deepfakes, solely on individuals.[394] Our previous report on media literacy emphasised that while attribution measures can help to empower users, platforms still have a responsibility to reduce users' exposure to misleading or harmful material in the first place.[395]

194. Labelling can also address the competition concerns highlighted by Ms Ganley by drawing a clear distinction between original, human-created works and AI-generated outputs. A research study from the Swiss Federal Institute of Intellectual Property observed "a growing appreciation and premium for human-created content", and noted that the "dichotomy between AI-generated and human-created content is shaping consumer preferences and behaviours".[396] In its report, *Rebooting Copyright: How the UK Can Be a Global Leader in the Arts and AI*, the Tony Blair Institute for Global Change said that preserving this distinction is "essential for fair competition".[397]

---

389 Written evidence from Google (AIC0010). The UK Government defines misinformation as "the inadvertent spread of false information", and disinformation as the "deliberate creation and spreading of false and/or manipulated information that is intended to deceive and mislead people, either for the purposes of causing harm, or for political, personal or financial gain". HM Government, *Fact Sheet on the CDU and RRU*, 9 June 2023

390 Written evidence from OpenAI (AIC0004) and Microsoft (AIC0012). OpenAI said that every video generated with its Sora model includes both visible and invisible provenance signals. Microsoft said it attaches signed provenance metadata to images generated with OpenAI's DALL-E 3 model in Azure OpenAI Service, Microsoft Designer and Paint, and that LinkedIn now automatically ingests and displays this information so that users can see whether an image has been generated or modified with AI.

391 Ofcom, *Deepfake Defences 2: The Attribution Toolkit*, p 26

392 Written evidence from Google (AIC0010)

393 Written evidence from techUK (AIC0011)

394 Ofcom, *Deepfake Defences 2: The Attribution Toolkit*, p 34

395 Communications and Digital Committee, *Media literacy* (3rd Report, Session 2024–26, HL Paper 163), para 85

396 Swiss Federal Institute of Intellectual Property, *Study on the Incentive Effects of Copyright for Generative Artificial Intelligence on Various Stakeholders*, October 2024, p 6

397 Tony Blair Institute for Global Change, *Rebooting Copyright: How the UK Can Be a Global Leader in the Arts and AI*, 2 April 2025

Legislation and regulation

195. The Independent Society of Musicians called for new legislation that would require mandatory labelling of AI-generated content "in all commercial contexts", supported by platform-level duties to detect, preserve and display provenance metadata.[398] The British Copyright Council pointed to the EU AI Act, noting that Article 50 requires certain AI outputs to be "marked in a machine-readable format", "detectable as artificially generated or manipulated", and based on "effective, interoperable, and robust technical solutions". It suggested that the UK could consider a comparable approach.[399]

196. We note that South Korea has recently brought into force a new law to introduce nationwide labelling obligations for AI-generated and deepfake content. Under this regime, providers of AI services must inform users when content has been produced using AI, using visible logos or explanatory text for content presented within a platform, and perceptible disclosures or metadata for content that can be downloaded and shared externally.[400]

197. There were diverging opinions, however, on whether such requirements should extend to *all* works that have been created using AI as a tool (commonly termed "AI-assisted" works), or just to wholly AI-generated content.[401] Stakeholders also differed on who should bear the obligation. The British Copyright Council argued that it should apply to "any person making the outputs available to consumers",[402] whereas Isabelle Doran, Chief Executive Officer of the Association of Photographers, said that the obligation should fall primarily on major platforms and AI providers.[403]

198. Other witnesses generally urged caution about legislating detailed technical requirements for rights reservation and output labelling at this stage. Google argued that, while governments may be tempted to promote a "one-size-fits-all" standard, it is important that industry retains "flexibility and room for the evolution of standards and best practices".[404] Meta similarly argued that prescribing specific approaches "would risk freezing in time the current state of development while new standards and new tools are still emerging, locking the UK out as the rest of the world continues to innovate".[405] Microsoft said that any legislative approach should recognise that provenance tools do not work equally well for all types of content, remain "technology agnostic", and avoid mandating requirements in areas "where they are not yet feasible or effective".[406]

199. ERA similarly suggested that data standards should be agreed "at sector level, between relevant stakeholders", with the Government's role being to "support the sector-led development of appropriate data standards and

398 Written evidence from Independent Society of Musicians (AIC0014)
399 For example, by amending the Consumer Protection Act 2015 to introduce labelling requirements for AI-generated content or by including such provisions in a dedicated AI framework. Written evidence from British Copyright Council (AIC0006)
400 Korea JoongAng Daily, *Could you be fined for AI content? What to know about Korea's latest technology law*, 25 January 2026
401 Written evidence from Alliance for Intellectual Property (AIC0002) and British Copyright Council (AIC0006)
402 Written evidence from British Copyright Council (AIC0006)
403 Q 11 (Isabelle Doran)
404 Written evidence from Google (AIC0010)
405 Written evidence from Meta (AIC0009)
406 Written evidence from Microsoft (AIC0012)

common practice".[407] We heard from one creator who is already experimenting with such approaches. Sam Hofman, a commercial still-life photographer, described developing workflows that maintain a timestamped record of creative decisions at each stage of an AI-assisted project. He called for an 'EXIF'-equivalent authorship metadata standard for AI-generated media—referring to the technical metadata embedded in digital photographs—so that a "structured, portable, verifiable record of the human creative process" can be embedded in or attached to the output file itself.[408]

200. Academic witnesses also cautioned against prescriptive regulation of specific watermarking or labelling technologies.[409] CREATe suggested that a more promising approach would be to prioritise content whose origins can be independently checked ("verified provenance"), and to shift more of the responsibility for verification onto platforms and intermediaries, "reversing the burden of proof". One possible legal mechanism, it argued, would be to explore liability privileges or other incentives for authenticated content, encouraging services to give preference to material that can be authenticated.[410]

201. Prof Collomosse similarly recommended that the Government could play a central role in "endorsing open standards", encouraging platforms to preserve provenance and labelling signals and recognising developers who respect them, for example through a kitemark-style scheme.[411] He also suggested that there could be a role for regulators in giving "guidance on best practice in industry at any given time".[412]

202. **Robust, visible labelling of AI-generated content must form a core part of the Government's approach to AI and copyright. Effective labelling can help preserve the integrity and value of human creativity, enable consumers to understand the content they encounter online, and support fair competition in creative markets which aim to maintain a premium on human-made work. These labels are, however, only one element of a wider response. They cannot be treated as a substitute for broader duties on platforms to limit users' exposure to misleading or harmful content and for sustained investment in effective media literacy interventions.**

203. ***The Government should support the development and adoption of open, interoperable and globally aligned standards for rights reservation, provenance and labelling, led by industry and the technical community, whilst recognising that voluntary approaches alone may be insufficient to ensure consistent and meaningful labelling of AI-generated content, including deepfakes. The Government should consider bringing forward legislation to place key duties on AI developers, service providers and online platforms to ensure that wholly AI-generated content is appropriately labelled when made available to users. Any such legislative framework should avoid locking in specific technical solutions that may quickly become outdated. Regulators should be empowered to issue guidance***

---

407 Written evidence from ERA (AIC0020)
408 Written evidence from Sam Hofman (AIC0028)
409 Q 103 (Prof John Collomosse) and written evidence from CREATe Centre, University of Glasgow (AIC0007)
410 Written evidence from CREATe Centre, University of Glasgow (AIC0007)
411 QQ 83, 103 (Prof John Collomosse)
412 Q 103 (Prof John Collomosse)

on best practice, encourage platforms to preserve provenance signals, and consider incentives for services that give prominence to authenticated content.

204. ***The Government should ensure that support for provenance and labelling technologies is matched by clear, enforceable expectations on major platforms and AI providers to implement media literacy initiatives that help users understand what AI-related labels can and cannot tell them about content. It should set out in its response to this report how it will work with Ofcom to embed these expectations within the wider online safety and media literacy regimes, and how platforms will be held to account for meeting them.***

CHAPTER 5: LICENSING

205. In the last two chapters, we have examined the transparency, rights-reservation and labelling arrangements that are needed to underpin a licensing-first approach. In this chapter, we examine the UK's existing licensing infrastructure and the Government's role in supporting an AI licensing market to thrive.

Licensing and the creative industries

206. A copyright licence is a legal agreement through which a copyright owner (the licensor) grants permission to another party (the licensee) to engage in activities that would otherwise infringe copyright under the CDPA, for example reproduction or distribution. The Creative Industries Policy and Evidence Centre (Creative PEC) described licensing as "the standard industry-led model by which rightsholders can exercise and exploit their rights."[413]

207. Licensing in the UK can be negotiated directly between individual rightsholders and users. Alternatively, it can be collectively administered through membership-based collective management organisations (CMOs). These are authorised by rightsholders to license, collect and distribute royalties on their behalf, often through blanket licences.[414] As the Intellectual Property Office notes, "the collective rights management landscape in the UK is well-developed and covers various sectors of the creative industries, including music, literature, art, and film."[415]

An emerging AI licensing market

208. Our evidence underlined that a market for licensing content for use by AI systems is already emerging. The Music Publishers Association pointed to a number of recent deals between AI developers and music companies,[416] for example, those between Udio and Universal Music Group,[417] and Suno and Warner Music Group.[418] Other witnesses identified deals that had been agreed with organisations including publishers, database providers, image libraries and media companies.[419] The Centre for Regulation of the Creative Economy (CREATe) shared a tracker of "known commercial agreements" between rightsholders and AI developers globally. Per its latest update in December 2025, 120 agreements had been recorded.[420]

209. CREATe told us: "Licensing agreements may be motivated by legal risk, access to quality content and data, the integration of AI into search, as well as technical convenience".[421] Several other witnesses emphasised that developers depend on high-quality, human-made content to sustain and improve their models. This is particularly the case for model fine-tuning, which requires

413 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)
414 Intellectual Property Office, *Licensing bodies and collective management organisations*, 23 January 2024
415 Intellectual Property Office, *Collective Rights Management in the UK: 2024–25*, 29 October 2025
416 Written evidence from Music Publishers Association (AIC0017)
417 Universal Music Group, *Universal Music Group and Udio announce Udio's first strategic agreements for new licensed AI music creation platform*, 29 October 2025
418 Warner Music Group, *Press release:* Warner Music Group and Suno forge groundbreaking partnership, 25 November 2025
419 QQ 3 (Anna Ganley), 8 (Isabelle Doran), 37 (Ed Newton-Rex), written evidence from OpenAI (AIC0004) and written evidence from Microsoft (AIC0012)
420 CREATe Centre, *The AI licensing economy*, 24 February 2025
421 Written evidence from CREATe Centre, University of Glasgow (AIC0007)

high-quality examples to enhance the accuracy of the generated outputs.[422] Since scraped data is often inaccurate and of poor quality,[423] most licensing deals to date appear to have been signed to access high-quality data for fine-tuning and retrieval augmented generation (RAG).[424] According to the Rt Hon Lisa Nandy MP, Secretary of State for Culture, Media and Sport, AI developers' focus was previously "very much about the quantum of data, but it is now about the quality of data."[425] The Creative Rights in AI Coalition told us that, as such, "the UK has an immense opportunity to exploit its wealth of high-quality data in the age of AI."[426]

Opportunities for a UK licensing market

210. Overall, we share the view put forward by the Alliance for Intellectual Property that:

> "There are significant opportunities for the UK to become a world-leading home of 'responsible' AI development, where developers use licensed content that enables their services and products to be deployed at scale, without questions over legal liability."[427]

211. As Antony Walker, Deputy Chief Executive Officer of techUK, highlighted, however, the majority of the deals announced to date have been direct licensing agreements between US firms, rather than UK-based businesses and rightsholders.[428] Technology sector stakeholders suggested that the introduction of a new commercial TDM exception would drive further licensing deals in the UK, by clarifying the instances in which a licence is required, and by encouraging a greater amount of model training to happen domestically.[429]

212. We were not persuaded on this point. techUK told us that already "structured licensing agreements are emerging between AI developers and content holders to facilitate data access to restricted data, such as paywalled and archived content"—that is, the kind of content that would in any event fall outside a broad TDM exception.[430] OpenAI noted that agreements such as these "will continue to come naturally and expand as consumer use of AI products grows."[431] As Roxanne Carter, Global IP Lead, Government Affairs and Public Policy at Google, explained, "when it comes to training AI models on content that is freely available on the open web, we do not believe that we should license that content", but "we are seeing a market develop for access to … archived content, for specialised datasets, for content

---

422 QQ 2 (Owen Meredith), 41 (Reema Selhi), 68 (Antony Walker), 126 (Roxanne Carter), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from OpenAI (AIC0004), written evidence from Google (AIC0010), written evidence from techUK (AIC0011), written evidence from Directors UK and British Equity Collecting Society (AIC0018) and supplementary written evidence from News Media Association (AIC0022)

423 See, for example, Perełkiewicz M, Poświata R, 'A review of the challenges with massive web-mined corpora used in large language models pre-training', International Conference on Artificial Intelligence and Soft Computing (June 2024), pp 153–63: https://doi.org/10.48550/arXiv.2407.07630

424 CREATe Centre, *The AI licensing economy*, 24 February 2025

425 Q 151 (Secretary of State for Science, Innovation and Technology)

426 Written evidence from Creative Rights in AI Coalition (AIC0001)

427 Written evidence from Alliance for Intellectual Property (AIC0002)

428 Q 58 (Antony Walker)

429 QQ 68 (Antony Walker), 107, 128, 134 (Roxanne Carter), written evidence from Google (AIC0010) and written evidence from techUK (AIC0011)

430 Written evidence from techUK (AIC0011)

431 Written evidence from OpenAI (AIC0004)

that may be off-platform in some way or for opted-out content."[432] It appears, therefore, that a broader TDM exception would only formalise the position that many AI developers have taken to date, rather than creating the wider opportunities for licensing of copyrighted material that rightsholders seek.

213. We do, however, recognise the points put forward by techUK that the current approach to licensing, where agreements have typically been negotiated bilaterally between large individual developers and publishers, "is not scalable". It pointed to barriers faced by "smaller developers that are unable to meet the licensing demands of large content organisations". In addition, the market is uneven: "content providers offering smaller pools of training data than the largest publishing groups are often overlooked by model providers."[433] Evidence from Microsoft highlighted why scale is a practical constraint: it told us that high-performing models are trained on the equivalent of tens of billions of news articles and that most copyrighted material is not actively managed, making it difficult to "identify and get licenses from the rightsholder".[434] Matthew Sinclair, Senior Director at the Computer and Communications Industry Association, emphasised that these practical barriers are intensified by the structural features of copyright as an automatic, unregistered right.[435]

214. Reema Selhi, Head of Policy and International at the Design and Artists Copyright Society (DACS), cautioned that the growing number of high-profile 'licensing' announcements cited were typically "content deals rather than copyright licences." These, she explained, provide access to material such as images and metadata, but "this is not the same as a copyright licence that rightsholders benefit from".[436] Serena Dedering, General Counsel and Company Secretary at the Copyright Licensing Agency, told us that due to the scale of the companies involved in deals to date, "there is a wide group of creators and rightsholders who are not yet able to access that market as it currently stands."[437]

215. Witnesses suggested that this was not due a lack of willingness from creative sector stakeholders. Ed Newton-Rex, Chief Executive Officer of Fairly Trained, told us that "there is a lot of appetite to license from rightsholders".[438] Ms Selhi told us that members of the visual arts CMO DACS "absolutely do want to license". She explained that in a survey of artists, 84% of respondents said that they would agree to license their works for AI training if they were remunerated, though many expressed concern about the previous use of their material without consent.[439]

216. Rightsholder representatives suggested that the barrier was instead that AI companies were not willing to engage.[440] Isabelle Doran, Chief Executive Officer of the Association of Photographers, told us: "One of the biggest challenges we have had, and we have seen it time and again, is the fact that

---

432 Q 107 (Roxanne Carter)
433 Written evidence from techUK (AIC0011)
434 Written evidence from Microsoft (AIC0012)
435 Q 60 (Matthew Sinclair). See also para 98.
436 Q 36 (Reema Selhi)
437 Q 35 (Serena Dederding)
438 Q 36 (Ed Newton-Rex)
439 Q 38 (Reema Selhi)
440 QQ 3 (Anna Ganley), 36–37 (Ed Newton-Rex) and written evidence from Alliance for Intellectual Property (AIC0002)

big tech will not come to the table."[441] At the same time, Vinous Ali, Deputy Executive Director at the Startup Coalition, reported that some smaller rightsholders have been reluctant to enter into licensing deals with startups, "because they are holding out for deals with larger language models."[442]

217. However, securing deals with large overseas AI companies may not be where the greatest opportunities lie for the growth of the UK's AI sector or for a thriving domestic licensing market. As noted in Chapter 2, training of large-scale generative AI models generally takes place outside the UK.[443] Evidence to our inquiry on scaleups suggested that the development of more specialised, fine-tuned models was where the UK's AI sector had particular potential to succeed.[444] Ms Ali echoed this, describing AI systems built for particular tasks or products, rather than large general-purpose models, as "where the UK might win in the AI race".[445] Several witnesses highlighted that such models present significant opportunities for content licensing.[446] As Creative PEC put it:

> "With large-scale AI generators for text and images being widely available and hosted in the US, licensing high-quality creative content for smaller, specialised generative AI models could be an opportunity for growth."[447]

Collective licensing and diverse routes to market

218. In our view, CMOs have a key part to play in establishing a UK licensing market that is accessible to players of all sizes. Ms Doran told us that collective rights management is "likely to be the best opportunity for individual creators" for whom direct licensing may not be an option.[448] techUK said that "collective licensing schemes … may have a role in improving content aggregation and remuneration particularly for smaller rightsholders", as well as in "providing a more easily accessible pool of data for smaller model developers".[449]

219. Our evidence emphasised that the UK already has a "very strong"[450] and "well-established"[451] collective licensing infrastructure, and that these organisations have extensive experience of handling multi-party rights and complex revenue-sharing models.[452] In addition, Ms Selhi told us that the collective rights management system "has always been developing new licences whenever new technologies have changed".[453] The CLA similarly stressed that over time it has "developed licensing solutions to meet evolving

441 Q 3 (Isabelle Doran)
442 Q 77 (Vinous Ali)
443 See para 18.
444 Communications and Digital Committee, *AI and creative technology scaleups: less talk, more action* (2nd Report, Session 2024–26, HL Paper 71), paras 99–102
445 Q 77 (Vinous Ali)
446 QQ 38 (Serena Dederding), 41 (Reema Selhi), 68 (Antony Walker)
447 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)
448 Q 8 (Isabelle Doran)
449 Written evidence from techUK (AIC0011)
450 Q 8 (Tom Kiehl)
451 Written evidence from Independent Society of Musicians (AIC0014)
452 QQ 3, 8, 11 (Isabelle Doran), 4 (Tom Kiehl), 42 (Serena Dederding, Reema Selhi), written evidence from Copyright Licensing Agency (AIC0005), written evidence from Independent Society of Musicians (AIC0014) and written evidence from Directors UK and British Equity Collecting Society (AIC0018)
453 Q 42 (Reema Selhi)

user needs in a changing technological landscape".[454] The Alliance for Intellectual Property told us that the UK "is world-leading in developing tailored licences for different use scenarios that support the whole creative ecosystem".[455]

220. In 2024, the CLA launched a licence that permits TDM for commercial purposes. It intends to launch a further licence for training and fine-tuning of AI systems, and RAG processes, in the next year in collaboration with two of its members, the Authors Licensing and Collecting Society and the Publishers' Licensing Services.[456]

221. The ISM felt that collective licensing should "form the backbone of the market", but suggested that "rightsholders should also retain the option to offer bespoke or direct licences where appropriate".[457] Ms Dederding agreed that collective and direct licensing could "complement" each other and "work together very well".[458] Mr Newton-Rex took the view that "any kind of licensing is workable; the key is having licensing in the first place".[459] techUK commented:

> "The market is already evolving to accommodate different needs, with a range of licensing models—including direct agreements, broad terms and conditions, and structured collective licences—demonstrating flexibility and adaptability, with ongoing discussions about how these models will evolve and scale."[460]

Ensuring that licensing benefits individual creators and performers

222. While many witnesses emphasised the strengths of the UK's existing licensing infrastructure, others cautioned that existing models may not ensure that individual creators, or certain creative subsectors, receive a fair share of the value generated by AI uses.[461] The Society of Authors (SoA) told us that AI training rights should be treated as 'new' rights that "are the authors to license", and that are not automatically granted to publishers. It proposed that the publisher's share of such rights should be "akin to that of an agent's commission for licensing rights", to ensure "fair remuneration" for authors.[462] The SoA also cited cases where authors had been opted into AI licensing agreements, without being informed of how, or by whom, their works had been used. It argued that publishers should be required to seek informed consent and to set out income percentages transparently if collective licensing for AI is to benefit authors in practice.

223. Concerns about bargaining power and contractual override were echoed in evidence from the Creative PEC. It warned that, in negotiations with large, international AI firms, copyright holders may "retain little control over their terms and conditions, with the potential for legal rights being overwritten

454 Written evidence from Copyright Licensing Agency (AIC0005)
455 Written evidence from Alliance for Intellectual Property (AIC0002)
456 Written evidence from Copyright Licensing Agency (AIC0005)
457 Written evidence from Independent Society of Musicians (AIC0014)
458 Q 41 (Serena Dederding)
459 Q 41 (Ed Newton-Rex)
460 Written evidence from techUK (AIC0011)
461 Q 20 (Dr Alina Trapova), written evidence from CREATe Centre, University of Glasgow (AIC0007), written evidence from Creative Industries Policy and Evidence Centre (AIC0008), written evidence from Directors UK and British Equity Collecting Society (AIC0018) and supplementary written evidence from Society of Authors (AIC0003)
462 Supplementary written evidence from Society of Authors (AIC0003)

though private contracting".[463] It suggested that licensing agreements should explicitly acknowledge the rights of copyright holders, and drew attention to proposals for a new specific "AI right" as one possible way to help creators negotiate returns on the use of their copyrighted works.[464]

224. The CMOs Directors UK and BECS expressed doubts that existing licensing structures would serve the interests of directors and audiovisual performers. Creators involved in films and TV series are often required to "assign all of their exclusive rights to the producer or distributor", who typically enjoys far greater bargaining power. They argued that individual directors and performers are unlikely to benefit directly from AI licensing deals concluded at that level. To address this, Directors UK and BECS called for legislative intervention, ideally through "the introduction of an unwaivable right to equitable remuneration for individual human creators, payable by AI developers and subject to mandatory collective management by CMOs".[465]

225. Witnesses also warned that some AI-related uses fall outside current copyright-based licensing frameworks altogether, particularly where performers' voices and likenesses are concerned. The ISM told us that licensing for the use of music, lyrics and recordings for AI training must be distinguished from the separate need to license the use of performers' voices and likenesses in AI systems.[466] As noted in Chapter 2,[467] the UK currently lacks personality rights, meaning that there is "no clear legal basis for licensing voice cloning".[468] The ISM argued that this gap must be addressed in parallel with music licensing solutions if performers are to retain meaningful control over AI uses of their identity. More broadly, the ISM proposed that any collective licensing arrangements for AI should be built around a "creator-first" remuneration model. This would require minimum rates for training uses, strong transparency and audit rights so creators can verify how their works have been used and how royalties have been calculated, and the use of existing CMO infrastructure to administer payments and manage multi-party splits effectively.

226. **A market for licensing content for use by AI systems is already emerging, as evidenced by recent deals struck between some of the largest AI and creative businesses. The UK has a strong foundation for AI licensing, with a wealth of creative content to draw on and a well-developed collective rights management infrastructure. This gives the UK a genuine opportunity to position itself as a global leader for responsible AI development based on licensed use of data.**

227. **A healthy UK AI licensing ecosystem should offer a variety of licensing models. Critically, it must be accessible to rightsholders and AI developers of different sizes, rather than being confined to transactions between the largest catalogues and technology firms. It is promising to see that stakeholder organisations are already working to bring new licensing schemes to market. Collective management organisations (CMOs) have an important role to play in ensuring**

---

463 Written evidence from Creative Industries Policy and Evidence Centre (AIC0008)
464 Creative Industries Policy and Evidence Centre, *Copyright and AI—a new AI Intellectual Property Right for composers, authors and artists*, 10 February 2025
465 Written evidence from Directors UK and British Equity Collecting Society (AIC0018)
466 Written evidence from Independent Society of Musicians (AIC0014)
467 See paras 72–77.
468 Written evidence from Independent Society of Musicians (AIC0014)

that individual creators can negotiate and benefit from AI licensing agreements.

228. ***As the UK AI licensing market develops, the Government must ensure that there are appropriate mechanisms in place to enable value to reach individual creators and rightsholders. This could include exploring the introduction of an unwaivable right to equitable remuneration for AI uses of their works and performances as training inputs and, where appropriate, as outputs, subject to mandatory collective management, and supporting creator-first remuneration models with appropriate transparency and audit arrangements. The Government should also determine whether relevant CMOs have sufficient resource to bring forward new solutions, and provide assistance to those organisations as required.***

The role of the Government

229. The Secretary of State for Culture, Media and Sport told us that a clear view had emerged from the technical working groups coordinated by the Government that it "ought to focus on creating the right conditions and frameworks, so issues like transparency, not intervening overly in licensing deals, which industry is already coming together to reach."[469] Broadly, this was echoed in our evidence. techUK said that it supported "an industry-led approach to licensing".[470] Ms Selhi told us that "the licensing market for all kinds of traditional media uses has not required much government intervention so far, apart from the regulatory role that the Intellectual Property Office plays".[471] Owen Meredith, Chief Executive Officer of the News Media Association, said: "I do not necessarily think that the Government need to create the market; they need to create the conditions for the market to develop."[472]

230. We have set out elsewhere several steps that the Government should take to create those conditions. On licensing, however, creative industry stakeholders were clear that the most impactful thing the Government could do would be to rule out the introduction of a new TDM exception.[473] The Creative Rights in AI Coalition argued that "this will be key in supercharging an already growing licensing market."[474] The British Copyright Council told us that: "It is the strong view of our members that whilst the Government leaves the option of a copyright exception on the table, there is a disincentive for AI developers to strike licensing deals with UK rightsholders."[475]

The Creative Content Exchange (CCE)

231. In the *Creative Industries Sector Plan*, the Government committed to establishing a 'Creative Content Exchange' (CCE) as a "trusted marketplace for selling, buying, licensing, and enabling permitted access to digitised cultural and creative assets", explicitly including its use a source of training

469 Q 139 (Secretary of State for Culture, Media and Sport)
470 Written evidence from techUK (AIC0011)
471 Q 50 (Reema Selhi)
472 Q 8 (Owen Meredith)
473 QQ 16 (Owen Meredith), 41 (Serena Dederding), 50 (Reema Selhi, Ed Newton-Rex), written evidence from Creative Rights in AI Coalition (AIC0001), written evidence from British Copyright Council (AIC0006) and supplementary written evidence from News Media Association (AIC0022)
474 Written evidence from Creative Rights in AI Coalition (AIC0001)
475 Written evidence from British Copyright Council (AIC0006)

data, "facilitating the development of high-value AI models".[476] In their written evidence to our inquiry, the Government described the CCE as a means of providing access to data "locked away in our economy … for the benefit of our taxpayers and businesses".[477] A pilot of the scheme will involve cultural institutions including Historic England, Science Museum Group and The National Archives. It will explore "how their digitised content can be used by consumers, technology companies and AI developers, while respecting the rights of creators and copyright owners". This will include "testing a range of commercial models for licensing".[478] The Secretary of State for Culture, Media and Sport explained that the Government does not see the CCE as a solution to all the issues stakeholders have raised in relation to AI and copyright but as a "way of helping to bring that solution into being".[479]

Stakeholder views on the CCE

232. We heard a wide range of views on the proposed CCE. Several witnesses suggested that, if appropriately designed, it could play a constructive role in the licensing ecosystem.[480] The Alliance for Intellectual Property noted that rightsholders support the broad objective of increasing licensing by AI providers, and that the CCE "may have a role in collaborating with existing licensing schemes."[481] Meta welcomed the CCE's potential to facilitate "voluntary collective licensing of datasets that reflect the cultural context of the UK". It argued that, if designed with "flexibility and feasibility in mind", it could "foster innovation" and "enhance UK competitiveness in the AI sector".[482] The Independent Society of Musicians described the CCE as a potentially "useful national infrastructure for rights reservation, transparency and licensing administration", but only under "strict conditions that protect creators", including integration with CMOs and strong auditing and governance to verify training uses and licensing compliance.[483]

233. At the same time, there was a strong consensus among creative-industry representatives that the CCE must not displace or undermine established licensing models.[484] The Creative Rights in AI Coalition added that the CCE would not be an appropriate marketplace for all creative subsectors or licensing scenarios, and that "rightsholders must remain free to license their works in whatever manner they see fit".[485]

234. Other witnesses questioned whether the CCE was needed at all.[486] Ms Selhi told us that the licensing market for traditional media "has not required much government intervention so far", and that the Government "need not necessarily play any greater role in the licensing market itself".[487]

---

476 HM Government, *Creative Industries Sector Plan*, 23 June 2025, p 17
477 Written evidence from HM Government (AIC0013)
478 Department for Science, Innovation and Technology, *Press release:* Targeted energy bill support and simpler access to legal guidance among plans to put data to work to improve lives, 26 January 2026
479 Q 153 (Secretary of State for Culture, Media and Sport)
480 Written evidence from Creative Rights in AI Coalition (AIC0001), Meta (AIC0009) and Independent Society of Musicians (AIC0014)
481 Written evidence from Alliance for Intellectual Property (AIC0002)
482 Written evidence from Meta (AIC0009)
483 Written evidence from Independent Society of Musicians (AIC0014)
484 Written evidence from Creative Rights in AI Coalition (AIC0001), Alliance for Intellectual Property (AIC0002) and Copyright Licensing Agency (AIC0005)
485 Written evidence from Creative Rights in AI Coalition (AIC0001)
486 Q 50 (Reema Selhi, Ed Newton-Rex)
487 Q 50 (Reema Selhi)

235. A further concern was that the CCE could ultimately be used to justify or facilitate a de facto opt-out regime for AI training. The ISM cautioned that the CCE "must not function as an opt-out mechanism or weaken existing rights", and should instead enhance creators' ability to exercise consent, enforce their rights and participate in a fair licensing market.[488] Mr Newton-Rex described the CCE as "a solution in search of a problem and a total waste of time", and said that he had "significant concerns that it will be used as a vehicle to allow a new text data mining exception". He suggested that it was "easy" to envisage the Government later announcing that AI companies would be allowed to "train on whatever they like", while pointing to the CCE as evidence of its support for licensing—an approach he characterised as, at best, "misguided" and, at worst, "intentionally distracting from the issues".[489]

236. Creative-sector witnesses emphasised, above all, that any licensing initiative such as the CCE could not substitute for clear legal rules and effective transparency and enforcement mechanisms.[490] The Creative Rights in AI Coalition said the "Government's first priority must be creating the right market conditions, rather than creating a marketplace", warning that "to focus on the latter without fixing the former would be to put the cart before the horse."[491]

237. **Given the UK's wealth of high-quality creative and informational data, AI firms will have strong incentives to negotiate for licensed access if Government policy creates clear incentives for a licensing-first approach, supported by effective transparency requirements and enforcement, rather than tolerating large-scale unlicensed training as a competitive advantage.**

238. **The Creative Content Exchange is a welcome pilot to test how cultural and heritage datasets can be licensed and used for AI and other data-driven purposes. However, it cannot by itself resolve underlying questions about compensation for the use of copyrighted works in AI training. Its value will depend on forming part of a wider framework that upholds lawful licensed use, strengthens transparency and enforcement, and ensures fair remuneration for rightsholders.**

239. ***The Government should focus its interventions on creating the right conditions to support the development of a UK AI licensing market, rather than relying on a single marketplace initiative. Any marketplace it establishes or supports should be designed to meet the needs of individual rightsholders and small and medium-sized UK AI firms, as well as larger players, and to complement existing licensing models instead of displacing them.***

488 Written evidence from Independent Society of Musicians (AIC0014)
489 Q 50 (Ed Newton-Rex)
490 QQ 50–51 (Serena Dederding) and written evidence from Creative Rights in AI Coalition (AIC0001)
491 Written evidence from Creative Rights in AI Coalition (AIC0001)

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

Chapter 2: AI, the creative industries and copyright reform

1. The interpretation of UK copyright law is a matter for the courts and future cases may help to set precedents in this area. However, the consistent call from technology sector stakeholders for a new, broad commercial text and data mining (TDM) exception expressly to enable more AI model training to take place in the UK suggests that they do not regard large-scale commercial training on copyright-protected works as clearly covered by the existing exceptions. If they did, a commercial TDM exception would be unnecessary. (Paragraph 36)

2. On this basis, the main uncertainty for large AI developers appears to lie in the question of whether their current and proposed training practices would withstand legal challenge if tested in court. Support for a broad commercial TDM exception should therefore be understood as an attempt to lower that litigation risk by weakening the current level of copyright protection, rather than as a neutral exercise in clarifying the law. We are also not persuaded that expanding the existing non-commercial research exception to cover all "pre-market" research and development is either necessary or desirable. (Paragraph 37)

3. Under existing law, copyright is engaged whenever the whole or a substantial part of a protected work is copied, including by storing it in digital form, subject only to specific statutory exceptions. We believe that the large-scale making and processing of digital copies of protected works for model training may therefore be characterised as reproduction, regardless of whether trained models retain human-readable copies or are capable of generating novel outputs. In our view, the lawfulness of using copyrighted content for AI training must be assessed under ordinary copyright principles and clearly defined exceptions. We do not accept the view that the copying and processing of protected works during training should be characterised as 'learning'. (Paragraph 38)

4. *We therefore consider that the Government should rule out any reform of the Copyright, Designs and Patents Act that would remove the incentive to license copyrighted works for AI training, and should instead focus on strengthening licensing, transparency and enforcement within the existing framework.* (Paragraph 39)

5. The UK creative industries are an economic powerhouse with huge growth potential, and a vital part of the UK's soft power. Their success has been underpinned by what witnesses described as an internationally-respected "gold-standard" copyright framework. While generative AI may present opportunities for creative innovation and further growth where appropriate safeguards are in place, it also poses substantial material risks to the work and livelihoods of individual creators. The impact of these is already being felt. (Paragraph 56)

6. We recognise that, in relation to AI-generated outputs that directly reproduce or 'regurgitate' protected works, existing copyright doctrines may provide a route in principle to challenge infringement. In practice, however, it is often prohibitively difficult for rightsholders to detect, evidence and enforce such claims. Despite this, the courts may be the only viable means of recovering damages for technology companies' past use of content for training purposes. (Paragraph 57)

7. AI has the potential to contribute to the UK's future economic growth and there could be a legitimate interest in enabling more AI training to take place on UK soil. However, in 2024 the UK AI sector contributed around £12 billion to the UK economy, compared with £124 billion in gross value added from the creative industries in 2023. The evidence available to date has not convincingly demonstrated that the existing copyright framework is the primary constraint on AI investment, or that weakening it through new or broader exceptions for commercial AI training would, in itself, significantly expand the sector. (Paragraph 58)

8. At the same time, the creative industries are already experiencing tangible harms from the current deployment of generative AI. Given their proven economic and cultural value, and the lack of robust evidence on the likely benefits of copyright reform to the UK AI sector, we do not consider that the case has been made for introducing new exceptions for AI training or for undertaking wider reform of UK copyright law. We welcome the Government's forthcoming economic impact assessment as an essential basis for any future, evidence-led decisions on whether changes to copyright law are justified. It is urgent that a regime is now created to safeguard creators' livelihoods going forward, while harnessing the potential of AI for creativity and economic growth. (Paragraph 59)

9. The application of copyright law to the development and operation of AI models is a technically complex and contested area. The Government is therefore right to seek evidence and engage extensively before reaching a decision, and we very much welcome that it has "reset" its previous approach. However, prolonged uncertainty and mixed public messaging on AI and copyright since this Government came into office have undermined trust and stalled both licensing and investment. (Paragraph 68)

10. The Government must accept that there is no solution that will satisfy all parties and that delaying in the hope of finding one risks exacerbating these issues further. Committing to a clear, stable policy framework, rooted in the UK's longstanding copyright principles and in the interests of its domestic creative and AI sectors, is increasingly urgent. (Paragraph 69)

11. *The Government should reach and publish a final, evidence-based decision on its approach to AI and copyright in the next 12 months. This decision should make clear that strong copyright protection and fair licensing for UK rightsholders are the default, and that policy choices will prioritise the long-term health of the UK's creative industries and domestic AI sector, rather than the interests of large multinational technology firms. In its March update published under the Data (Use and Access) Act, the Government should set out the concrete steps and timetable it will follow to reach this decision, and detail any further evidence, analysis, or engagement it requires.* (Paragraph 70)

12. *In the meantime, we urge the Government to issue a clear public statement setting out an expectation that commercial AI developers operating in the UK should obtain appropriate licences when using copyrighted works to train generative AI models. That statement should signal the Government's support for a licensing-first approach as the baseline for the UK market, and should be published at the earliest opportunity.* (Paragraph 71)

13. The absence in UK law of a robust personality right or specific protection for digital likeness means that creators and performers lack an adequate

basis to challenge harmful outputs that imitate their distinctive style, voice or persona without reproducing a specific underlying work. (Paragraph 83)

14. *The Government should introduce protections against unauthorised digital replicas and 'in the style of' uses. Any new framework should give creators and performers clear, enforceable control over the commercial exploitation of their identity, while appropriately safeguarding freedom of expression and other legitimate uses.* (Paragraph 84)

Chapter 3: Transparency

15. Meaningful transparency reporting on AI training data and data use practices is a prerequisite for rebuilding trust between the AI and creative sectors and for enabling effective licensing and enforcement. Voluntary codes alone are unlikely to drive the level of compliance required to address rightsholders' concerns and create a level playing field for developers. (Paragraph 114)

16. Our evidence was clear that high-level, aggregate transparency disclosures are unlikely to meet rightsholders' needs and that more granular transparency reporting is required. However, in their evidence to us, AI developers objected to the prospect of granular transparency reporting, highlighting technical barriers, commercial sensitivity, competitive dynamics and compliance costs. We are persuaded that reaching a solution that meets the needs of both sides is feasible. The Government has a clear role to play in facilitating discussions between AI firms and rightsholder representatives to develop proportionate, workable proposals. (Paragraph 115)

17. *The Government should seek to set best practice, working in collaboration with international partners, by developing a clear, proportionate framework for transparency reporting on data use in AI training. This should go beyond existing voluntary practice and improve on the high-level summaries required under current EU provisions, while remaining mindful of the potential impact on the UK's AI ecosystem.* (Paragraph 116)

18. *Transparency obligations for large AI developers should be given statutory weight. The Government should identify the appropriate regulatory body to set standardised reporting formats, monitor compliance and take enforcement action as required. In designing such obligations, particular care should be taken to avoid disproportionate burdens on smaller UK-based firms, for example through tailored requirements.* (Paragraph 117)

19. *The Government should also explore the potential for requiring commercial AI developers to make more granular confidential disclosures about their training data and methods to the relevant regulator, along the lines of the model documentation form incorporated into the EU's General-Purpose AI Code of Practice, so as to balance developers' commercial interests with rightsholders' needs for meaningful transparency.* (Paragraph 118)

20. We heard compelling evidence that any UK transparency regime for AI should, in principle, apply to all models available on the UK market, regardless of where they are trained. At the same time, it is important to recognise the territorial nature of copyright and the limits of its ability to regulate training that occurs entirely overseas. (Paragraph 126)

21. *In giving effect to that principle, the Government should consider how public procurement and regulatory tools could support compliance with UK transparency requirements by AI developers operating in the UK. At a minimum, it should*

design transparency requirements in a way that minimises incentives for UK-based developers to relocate training abroad, or for frontier model providers to delay or withhold the release of new models in the UK, which could be detrimental to UK creators, innovators and consumers. (Paragraph 127)

22. International examples such as the Swiss AI Initiative's Apertus system and the Allen Institute for AI's OLMo project indicate that it is possible to build competitive models whose training data and development pipelines are open to external scrutiny. Any UK approach to sovereign AI should draw on these lessons and align future capability with robust standards of compliance, accountability and respect for copyright. (Paragraph 134)

23. *The UK should not resign itself to long-term dependence on opaque models trained overseas. Instead, the Government's sovereign AI efforts should prioritise the development and adoption of domestically governed models with transparency built in by design, including clear information about training data and development processes. The Government must be clear that where the use of copyrighted data in the training of such models is not covered by an existing exception, remuneration for rightsholders will be provided.* (Paragraph 135)

Chapter 4: Emerging technical solutions

24. We welcome the proposals from the Competition and Markets Authority (CMA) to introduce conduct requirements for Google's search services following its designation of Google with 'strategic market status'. For these measures to address the concerns raised in our inquiry, they must ensure publishers have meaningful technical control over whether their content is used for Google Search, for AI training, and for retrieval-augmented generation or 'grounding' features across Google's products, including through clearer signalling of distinct uses. We also note more broadly the importance of the CMA's digital markets regulatory regime and will continue to scrutinise the effectiveness of its implementation. (Paragraph 158)

25. A broad TDM exception with an opt-out rights-reservation mechanism, as envisaged under "option 3" of the Government's consultation, is not a viable foundation for the UK's AI and copyright regime. The tools currently available for rights reservation are fragmented, poorly understood and place an unreasonable burden on individual creators. (Paragraph 175)

26. Experience in the EU suggests that comparable provisions have neither delivered reliable control for rightsholders nor a robust licensing market. We welcome the Secretary of State's acknowledgement that there is currently no workable opt-out proposal on the table and consider that the Government should now draw a clear line under this approach. The UK's starting point should remain a licensing-first model in which permission is required to use copyright-protected works for AI training. (Paragraph 176)

27. *In the light of this, we urge the Government to follow the example of the Australian Government and make a clear public statement that it will not bring forward proposals for a new commercial TDM exception with an opt-out-based rights-reservation mechanism.* (Paragraph 177)

28. Effective, machine-readable rights-reservation mechanisms will be an essential component of any sustainable licensing regime. Such tools are needed to give rightsholders granular control over how their works are used by AI systems, and to enable developers to assemble the large, diverse

datasets they require while operating within the law. Existing site-level rights-reservation tools were not designed with modern AI use cases in mind and have significant limitations in this context. (Paragraph 186)

29. It is encouraging to see industry bringing forward new solutions that may help to address these shortcomings. Robust provenance technologies combining signed metadata, fingerprinting and watermarking seem particularly promising. By making consent and licensing signals durable as content moves through digital supply chains, these tools could help ensure rightsholders' preferences are respected, as well as creating new opportunities for remuneration at the point of data access. (Paragraph 187)

30. Robust, visible labelling of AI-generated content must form a core part of the Government's approach to AI and copyright. Effective labelling can help preserve the integrity and value of human creativity, enable consumers to understand the content they encounter online, and support fair competition in creative markets which aim to maintain a premium on human-made work. These labels are, however, only one element of a wider response. They cannot be treated as a substitute for broader duties on platforms to limit users' exposure to misleading or harmful content and for sustained investment in effective media literacy interventions. (Paragraph 202)

31. *The Government should support the development and adoption of open, interoperable and globally aligned standards for rights reservation, provenance and labelling, led by industry and the technical community, whilst recognising that voluntary approaches alone may be insufficient to ensure consistent and meaningful labelling of AI-generated content, including deepfakes. The Government should consider bringing forward legislation to place key duties on AI developers, service providers and online platforms to ensure that wholly AI-generated content is appropriately labelled when made available to users. Any such legislative framework should avoid locking in specific technical solutions that may quickly become outdated. Regulators should be empowered to issue guidance on best practice, encourage platforms to preserve provenance signals, and consider incentives for services that give prominence to authenticated content.* (Paragraph 203)

32. *The Government should ensure that support for provenance and labelling technologies is matched by clear, enforceable expectations on major platforms and AI providers to implement media literacy initiatives that help users understand what AI-related labels can and cannot tell them about content. It should set out in its response to this report how it will work with Ofcom to embed these expectations within the wider online safety and media literacy regimes, and how platforms will be held to account for meeting them.* (Paragraph 204)

Chapter 5: Licensing

33. A market for licensing content for use by AI systems is already emerging, as evidenced by recent deals struck between some of the largest AI and creative businesses. The UK has a strong foundation for AI licensing, with a wealth of creative content to draw on and a well-developed collective rights management infrastructure. This gives the UK a genuine opportunity to position itself as a global leader for responsible AI development based on licensed use of data. (Paragraph 226)

34. A healthy UK AI licensing ecosystem should offer a variety of licensing models. Critically, it must be accessible to rightsholders and AI developers of different sizes, rather than being confined to transactions between the largest

catalogues and technology firms. It is promising to see that stakeholder organisations are already working to bring new licensing schemes to market. Collective management organisations (CMOs) have an important role to play in ensuring that individual creators can negotiate and benefit from AI licensing agreements. (Paragraph 227)

35. *As the UK AI licensing market develops, the Government must ensure that there are appropriate mechanisms in place to enable value to reach individual creators and rightsholders. This could include exploring the introduction of an unwaivable right to equitable remuneration for AI uses of their works and performances as training inputs and, where appropriate, as outputs, subject to mandatory collective management, and supporting creator-first remuneration models with appropriate transparency and audit arrangements. The Government should also determine whether relevant CMOs have sufficient resource to bring forward new solutions, and provide assistance to those organisations as required.* (Paragraph 228)

36. Given the UK's wealth of high-quality creative and informational data, AI firms will have strong incentives to negotiate for licensed access if Government policy creates clear incentives for a licensing-first approach, supported by effective transparency requirements and enforcement, rather than tolerating large-scale unlicensed training as a competitive advantage. (Paragraph 237)

37. The Creative Content Exchange is a welcome pilot to test how cultural and heritage datasets can be licensed and used for AI and other data-driven purposes. However, it cannot by itself resolve underlying questions about compensation for the use of copyrighted works in AI training. Its value will depend on forming part of a wider framework that upholds lawful licensed use, strengthens transparency and enforcement, and ensures fair remuneration for rightsholders. (Paragraph 238)

38. *The Government should focus its interventions on creating the right conditions to support the development of a UK AI licensing market, rather than relying on a single marketplace initiative. Any marketplace it establishes or supports should be designed to meet the needs of individual rightsholders and small and medium-sized UK AI firms, as well as larger players, and to complement existing licensing models instead of displacing them.* (Paragraph 239)

APPENDIX 1: LIST OF MEMBERS AND DECLARATIONS OF INTEREST

Members

Baroness Caine of Kentish Town (from January 2026)
Viscount Colville of Culross
Baroness Elliott of Whitburn Bay (from July 2025)
Baroness Fleet
Baroness Healy of Primrose Hill (until January 2026)
Lord Holmes of Richmond
Baroness Keeley (Chair)
Lord Kirkhope of Harrogate (from January 2026)
Lord Knight of Weymouth
The Lord Bishop of Leeds (until November 2025)
Lord McNally
Baroness Owen of Alderley Edge
Lord Storey
Lord Tarassenko (from January 2026)
Baroness Wheatcroft (until January 2026)
The Lord Bishop of Winchester (from January 2026)

Declarations of interest

Baroness Caine of Kentish Town
No relevant interests declared
Viscount Colville of Culross
No relevant interests declared
Baroness Elliott of Whitburn Bay
No relevant interests declared
Baroness Fleet
No relevant interests declared
Baroness Healy of Primrose Hill
No relevant interests declared
Lord Holmes of Richmond
Adviser, Simmons and Simmons LLP (emerging technologies including AI)
Member, Global Advisory Board, Endava plc (IT)
Member, Technology and Science Advisory Committee, The Crown Estate
Adviser, Submer Technologies Ltd (data-centre technologies and IT)
Baroness Keeley (Chair)
No relevant interests declared
Lord Kirkhope of Harrogate
No relevant interests declared
Lord Knight of Weymouth
No relevant interests declared
The Lord Bishop of Leeds
No relevant interests declared
Lord McNally
No relevant interests declared
Baroness Owen of Alderley Edge
Guest of Google Future Forum, December 2024; transport, accommodation and subsistence costs met by Google
Lord Storey
No relevant interests declared

Lord Tarassenko

Research Professor, Department of Engineering Science, University of Oxford

Non-executive Director, LIO (formerly Oxehealth Limited)

Machine Learning Adviser, Deep Medical

Member of Advisory Board, Time Trace Labs

Senior Adviser, Alan Turing Institute

Baroness Wheatcroft

Chair, Financial Times Oversight and Complaints Committee

Regular columnist, 'The New World'

The Lord Bishop of Winchester

No relevant interests declared

A full list of Members' interests can be found in the Register of Lords' Interests: https://members.parliament.uk/members/lords/interests/register-of-lords-interests

APPENDIX 2: LIST OF WITNESSES

Evidence is published online at https://committees.parliament.uk/committee/170/communications-and-digital-committee/publications/

Evidence received by the Committee is listed below in alphabetical order.

Alphabetical list of witnesses

Alliance for Intellectual Property	AIC0002	
Association of Photographers		QQ 1–16, Isabelle Doran, Chief Executive Officer
Dr Hayleigh Bosher, Reader (Associate Professor) in Intellectual Property Law, Brunel, University of London	AIC0025	
British Copyright Council	AIC0006	
The British Equity Collecting Society (BECS)	AIC0018 AIC0019	
Charismatic.ai		QQ 106–34, Guy Gadney, Chief Executive Officer
Cloudflare	AIC0015	QQ 79–105, Ed Conolly, Vice President Engineering
Professor John Collomosse, Professor of Computer Vision and AI, University of Surrey		QQ 79–105
Computer and Communications Industry Association (CCIA)		QQ 56–78, Matthew Sinclair, Senior Director
The Copyright Licensing Agency	AIC0005	QQ 34–55, Serena Dederding, General Counsel and Company Secretary
CREATe Centre, University of Glasgow	AIC0007	
Creative Industries Policy and Evidence Centre (Creative PEC)	AIC0008	
Creative Rights in AI Coalition	AIC0001	
Design and Artists Copyright Society (DACS)		QQ 34–55, Reema Selhi, Head of Policy and International
Directors UK [joint submission]	AIC0018	
ERA (The Digital Entertainment and Retail Association)	AIC0020	

Fairly Trained		QQ 34–55, Ed Newton-Rex, Chief Executive Officer
Google	AIC0010 AIC0024	QQ 106–34, Roxanne Carter, Global IP Lead, Government Affairs and Public Policy
Dr Andres Guadamuz, Reader in Intellectual Property Law, University of Sussex	AIC0026	
HM Government—Department for Culture, Media and Sport [joint submission]	AIC0013	QQ 135–55, Rt Hon Lisa Nandy MP, Secretary of State for Culture, Media and Sport and Ruth Hannant, Director General for Society, Media and Culture
HM Government—Department for Science, Innovation and Technology [joint submission]	AIC0013	QQ 135–55, Rt Hon Liz Kendall MP, Secretary of State for Science, Innovation and Technology and Oliver Ilott, Interim Director General for Artificial Intelligence
HM Government—Intellectual Property Office [joint submission]	AIC0013	
Sam Hofman, Director, Hofman Studio	AIC0028	
Independent Society of Musicians	AIC0014	
Professor Sylvester Kaczmarek	AIC0016	
Professor Martin Kretschmer, Professor of Intellectual Property Law, University of Glasgow	AIC0027	
Meta	AIC0009	
Microsoft	AIC0012	
Music Publishers Association	AIC0017	
News Media Association	AIC0022 AIC0029	QQ 1–16, Owen Meredith, Chief Executive Officer
OpenAI	AIC0004	
ProRata.ai		QQ 79–105, Eugene Huang, Senior Strategy Advisor and Co-founder

Professor Eleonora Rosati, Professor of Intellectual Property Law, Stockholm University		QQ 17–33
RSL Collective	AIC0023	
Society of Authors	AIC0003	QQ 1–16, Anna Ganley, Chief Executive Officer
Startup Coalition		QQ 56–78, Vinous Ali, Deputy Executive Director
techUK	AIC0011	QQ 56–78, Antony Walker, Deputy Chief Executive Officer
Dr Alina Trapova, Lecturer in Intellectual Property Law, University College London	AIC0021	QQ 17–33
UK Music		QQ 1–16, Tom Kiehl, Chief Executive Officer