



HM Government

Report on Copyright and Artificial Intelligence



Department for Science, Innovation and Technology

Department for Culture, Media and Sport

Report on Copyright and Artificial Intelligence

Presented to Parliament pursuant to Section 136 of the Data (Use and Access)
Act 2025

March 2026



© Crown copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at www.gov.uk/official-documents.

Any enquiries regarding this publication should be sent to us at:

Department for Science, Innovation and Technology
22 Whitehall, London SW1A 2EG

ISBN 978-1-5286-6308-3
E03546782 03/26

Printed on paper containing 40% recycled fibre content minimum

Printed in the UK by HH Associates Ltd. on behalf of the Controller of His Majesty's Stationery Office

Contents

Section A: Report on Copyright and Artificial Intelligence	7
Overview	7
Copyright and artificial intelligence	8
The Copyright and AI Consultation	8
The government's approach	10
A copyright exception for AI training	11
Transparency over the content and data used to develop AI systems	11
Labelling of AI and human-created content	12
Technical tools and standards	12
Licensing	13
Enforcement	13
Computer-generated works	14
Digital replicas	14
Section B: Copyright and AI systems	15
AI system development	16
Copyright law	21
Types of protection	21
Rights and exceptions	22
Copyright in other countries	23
Copyright and artificial intelligence models	25
Status Quo: Views from consultation	32
Status Quo: Assessment	34
Section C: Copyright and AI systems – Consultation options	36
Overview	36
A broad data mining exception, which allows right holders to reserve their rights by opting out (consultation option 3)	36
Exception with opt-out: views from consultation	38
Opt-out exception: potential effects	40

No new exception, copyright is strengthened so licensing is required for AI development (consultation option 1).....	44
A broad data mining exception, with no opt-out (consultation option 2).....	50
Alternative approaches.....	54
Section D: Input transparency	60
Input transparency.....	60
Web crawler transparency	61
Legislative approaches to transparency.....	61
International approaches.....	62
Views from the consultation.....	64
Technical working groups.....	67
Conclusions and proposals	67
Section E: Output transparency	69
Overview.....	69
International approaches.....	70
Views from the consultation.....	72
Technical working groups.....	74
Conclusions and proposals.....	74
Section F: Technical tools and standards	75
Overview.....	75
Current challenges.....	79
Views from consultation.....	81
Conclusions and proposals.....	85
Section G: Licensing	86
The legal basis for copyright licensing.....	86
Current licensing environment for AI.....	88
Views from the consultation.....	92
Conclusions and proposals.....	95
Section H: Enforcement	96

Overview of existing mechanisms.....	96
The role of a regulator.....	99
Market access and extra-territoriality.....	100
Views from the consultation.....	101
Conclusions and proposals.....	102
Section I: Computer-generated works.....	104
Rationale for computer-generated works protection.....	104
Other forms of copyright protection for AI outputs.....	105
Legal scope of computer-generated works protection.....	105
International comparisons	106
Views from the consultation	108
Conclusions and proposals.....	111
Section J: Digital replicas.....	112
Overview of impacts from digital replicas	112
Application of existing UK law	113
Commercial-creative impacts of digital replicas.....	114
Existing UK intellectual property law	115
International law and developments.....	118
Views from the consultation.....	119
Stakeholder engagement.....	122
Conclusion and proposals	123



Section A: Report on Copyright and Artificial Intelligence

Overview

1. This report has been prepared pursuant to section 136 of the Data (Use and Access) Act 2025 (“the D(UA) Act”)¹. It considers the use of copyright works in the development of artificial intelligence (AI) systems. It follows the government’s consultation on Copyright and Artificial Intelligence, which ran from 17 December 2024 to 25 February 2025, as well as more recent engagement and developments.
2. The first section of the report considers the four policy options on copyright and the training of artificial intelligence models set out in the government’s consultation, as well as alternative approaches.
3. Subsequent sections consider aspects of copyright and AI training, including the wider regulatory environment, in more depth. They consider:
 - the effect of copyright on access to, and use of, data by AI developers;
 - transparency about the access to, and use of, copyright works by AI developers, and about the outputs of AI systems;
 - technical measures and standards that may be used to control the access to, and use of, copyright works to develop AI systems;
 - licensing of copyright works for AI development;
 - enforcement of requirements and restrictions relating to the access to, and use of, copyright works to develop AI systems, and relating to their outputs.
4. These correspond to the themes listed in section 136(3) of the D(UA) Act. The sections draw on consultation responses and other information, including insights from stakeholder roundtable meetings and technical working groups.
5. The report also considers the other areas explored in the Copyright and AI consultation, which are not covered by the requirements of the D(UA) Act – computer-generated works and digital replicas.
6. The report is published alongside an economic impact assessment, as required by section 135 of the D(UA) Act².

¹ Section 136 of the Data (Use and Access) Act can be accessed here: <https://www.legislation.gov.uk/ukpga/2025/18/section/136>

² Section 135 of the Data (Use and Access) Act can be accessed here: <https://www.legislation.gov.uk/ukpga/2025/18/section/135>

Copyright and artificial intelligence

7. The first copyright act (the 1710 Statute of Anne) gave authors, for the first time, a statutory right to control who could reproduce their books. By controlling the right to copy, an author could license their work to publishers in exchange for payment, enabling them to make a living. The Statute sought to encourage the writing of “useful books” by “learned men”.
8. Technology has given rise to new types of media, and new ways to copy and communicate creative works. Copyright developed over many years in response to technological change. Modern copyright law now gives creators the ability to control who may copy, distribute, and communicate their work through a wide range of channels and formats. Copyright enables many creators to make a living and is the foundation on which our creative industries are built.
9. Copyright law has long sought to balance the rights it gives to creators with the wider public benefit that comes from the exchange of information. Copyright protects the expression of creativity in a person’s work but not any facts contained in it. It is also subject to several exceptions that support free expression and the flow of information. These constraints on copyright help to ensure that knowledge is accessible and can benefit society.
10. Advances in AI, and particularly “generative” AI, pose new questions for copyright, and how to strike this balance. Frontier AI models require billions of inputs for training. These inputs are often copyright works. They learn from the information provided by those works, translating it into statistical representations of the world. But they would not be able to learn without human creativity, and their outputs may compete with the very creators they learn from.
11. We want to ensure that our approach to copyright allows us to realise the extraordinary potential of AI to grow the economy, create new, more rewarding jobs, and improve living standards, while protecting the UK’s position as a creative powerhouse.

The Copyright and AI Consultation

12. In December 2024 the government launched a consultation on copyright and artificial intelligence. We wanted to explore how copyright can best support these aims and what action, if any, needs to be taken. We received 11,520 responses: 10,110 via Citizen Space and 1,410 by email. Responses were received from a range of respondents, including creators and right holders, developers of AI models and applications, academics, researchers, cultural heritage organisations, and legal professionals. They included individuals, and micro-, small- and medium-sized enterprises (SMEs).³, and groups which represent them, as well as larger businesses and other organisations. Consideration has been given to all the consultation responses.

³ In this report, “SME” refers to micro (with fewer than 10 staff), small (at least 10 but fewer than 50 staff), and medium-sized (at least 50 but fewer than 250 staff) businesses.

13. The majority of responses reflected views of right holders and the creative industries. These included a significant number of responses from individuals, including individual creators and performers. Many of these responses either repeated or built upon a set of template letters, or template survey responses, that were created and distributed by interested organisations or individuals. A significant, but smaller, number of responses reflected views of the technology sector, including AI developers. Users of AI technology, including researchers and cultural heritage organisations, also responded to the consultation, but these may not reflect the full variety of sectors in which AI is deployed. Each response has been read individually by officials, and the views in those responses inform this report.
14. The consultation set out four options for copyright and AI policy. These were:
 - Do nothing (status quo): copyright and related laws remain as they are (option 0);
 - Strengthen copyright requiring licensing in all cases (option 1);
 - A broad data mining exception (option 2);
 - A data mining exception with opt-out and transparency measures (option 3).
15. Options 1 to 3 incorporated packages of measures, including one or more of:
 - Revisions to copyright law;
 - Measures promoting transparency about copyright works used to train AI models;
 - Measures promoting the use of technological tools and standards;
 - Measures on the licensing of copyright works.
16. While the consultation was open, copyright and AI became a key topic of debate during the passage of the D(UA) Bill. This resulted in the commitments in the D(UA) Act which this report fulfils.
17. The government's originally preferred consultation proposal – a broad exception with opt-out (option 3) – was rejected by most respondents to the consultation. Many from the creative industries were concerned that a broad exception would undermine the value of their work, and that an opt-out would be impractical. Concerns were also raised by some in the AI and research sectors who felt that this option would be more restrictive than the approach taken in other countries, so would not achieve the aim of making the UK internationally competitive for AI training and development.

The government's approach

18. We must take the time needed to get this right. We will not introduce reforms to copyright law until we are confident that they will meet our objectives for the economy and UK citizens. This means protecting the UK's position as a creative powerhouse, while unlocking the extraordinary potential of AI to grow the economy and improve lives. Any reform must ensure that right holders can be fairly rewarded for the economic value their work creates, and that they are protected against unlawful and unfair use of their work. It must also ensure that AI developers can access high quality content. It is clear through the consultation and our subsequent engagement that there is no consensus on how these objectives should be achieved.
19. We have limited and uncertain evidence on the impact of copyright on the development and deployment of AI in the UK. We must continue to build this evidence base. The economic impact assessment on copyright and AI considers the available evidence in more detail. There is uncertainty about the extent to which reforms to copyright would secure significant investment in AI development in the UK, and further uncertainty about the benefits to the security and prosperity of the UK economy from that greater investment. It is also unclear how copyright impacts the deployment of AI, particularly as the use of autonomous agents becomes more widespread. Finally, there is some ongoing uncertainty around how a broad exception would impact licensing markets, given most training is likely to occur in jurisdictions with the most permissive regimes, and such regimes are being litigated.
20. Meanwhile, there have been important developments in the wider market since the close of the consultation. The volume of litigation has increased, particularly in the USA. New transparency rules have come into effect – and are under review – in the EU and have been introduced in parts of the USA. Technical standards are also developing, including those on the use of web crawlers. The licensing market continues to grow to the benefit of right holders and AI developers, though limited public information on licensing deals makes it difficult to fully assess their impact. These developments will significantly shape the impact of reforms in the UK and it is important they inform our final policy position.
21. We propose to address the gaps in evidence on copyright reform, consider alternative options and review our approach in light of wider market and international developments. Alongside this, we propose to take steps to help right holders control and license their work, including through encouraging greater transparency. We also propose to explore options for supporting human creativity and artistry. Across this, we will continue to seek input from voices across the economy, and engage with Parliament and technical experts to ensure any reform drives growth and supports adoption and diffusion of artificial intelligence.
22. We recognise that these issues are of great interest to parliamentarians, and we have engaged with many of them through our parliamentary working groups. The House of Lords Communications and Digital Committee has undertaken its own inquiry into this issue, and we will carefully consider the findings of the committee and the evidence they have collated as we take this work forward.
23. The rest of this section sets out our approach on each of the areas set out in the D(UA) Act. We also set out our approach on computer-generated works and digital replicas, which were addressed in the same consultation. A full analysis of each of

these issues, and further detail of consultation responses, is provided in subsequent sections of this report.

24. Where this report proposes that further work be undertaken, or further consideration be given, that will include consideration of the likely effect of further proposals in the UK on copyright owners, AI developers and AI users, including individuals and SMEs. That work will also take account of AI systems developed both in the UK and elsewhere.

A copyright exception for AI training

25. The majority of respondents rejected the originally preferred proposal in our consultation: a broad exception with opt-out. Many responses were from the creative industries, who were concerned a broad exception would allow generative AI to learn from their works, without compensation, and in direct competition to them.
26. Meanwhile many AI developers and others from the technology sector said that exceptions to copyright would support AI innovation in the UK. Others noted that exceptions could support the use of AI in the wider economy, including universities, researchers and scientists.
27. In light of the strong views from the consultation, the gaps in evidence and the rapidly evolving AI sector and international context, a broad copyright exception with opt-out is no longer the government's preferred way forward. We propose to gather further evidence on how copyright laws are impacting the development and deployment of AI across the economy. We will consider and engage stakeholders on other potential policy approaches. We will also continue to monitor developments in technology, litigation, international approaches, and the licensing market.

Transparency over the content and data used to develop AI systems

28. Developers take different approaches to transparency over the content used to train their models. Some countries have introduced transparency regulations that require AI developers to disclose sources of training data, with the aim of supporting compliance with copyright law.
29. The majority of respondents to the consultation argued that AI developers should disclose the sources of their training material. There was strong support for mandatory standards on transparency from the creative industries, who saw it as key to enforcing their rights. While technology companies also supported transparency, they argued that commitments should be high-level and industry-led, to help ensure they were proportionate. Other countries have adopted transparency regulations, including the EU, which has transparency requirements as well as data mining exceptions and other measures. Views were mixed on whether we should follow the transparency approaches adopted by the EU and other countries.
30. We agree that greater transparency about how AI developers train their models, including the content and data they use, can help right holders assert their rights. We propose to continue monitoring the effects of transparency rules in other

countries and consider our approach in the UK. Our approach must promote clarity and enforcement for right holders, without disproportionate effects on AI development or deployment in the UK. We propose to work with industry and experts to develop best practice on input transparency, to help right holders assert their rights. This will inform any future potential legislation.

Labelling of AI and human-created content

31. Labelling content so it is clear whether it has been made using AI can inform people's choices. It may also help protect against disinformation and harmful deepfakes. There are no obligations in the UK for AI-generated content to be labelled, but many services already include labelling technology and several countries have introduced labelling rules.
32. There was relative consensus amongst respondents to the consultation in favour of the principle of labelling content that has been generated using AI. However, many noted the importance of AI to the creative process and therefore favoured labelling rules for wholly AI-generated content, and a more nuanced approach to AI-assisted content, as well as different approaches to different media types.
33. We propose to work with industry to explore best practice on labelling AI-generated content. We propose to continue monitoring international developments and to work with international partners to support the development of common solutions.

Technical tools and standards

34. Technical tools and standards enable right holders to express how their work can be accessed and used, helping them to control and license it. They play an important role in mediating the use of online content with AI, through web crawlers and AI agents.
35. The market for these tools and standards is developing rapidly. However, currently they do not support all right holders' needs, and there are challenges with adoption and compliance. Respondents from the AI, creative and other sectors support the development and use of technical tools and standards, but there were different views on the extent to which government intervention is needed.
36. We propose to keep the need for regulation of technical tools and standards under review, and to continue to monitor international developments. We propose to work with experts and industry to support best practice and adoption of market-led tools and standards.

Licensing

37. Creators' work is widely used to train AI models. Licensing is important to ensure creators are paid when their work is used and in incentivising new creative content. Licensing of works for use with AI, whether for training or other uses within the AI ecosystem, can also give creators opportunities to generate new sources of income.
38. Many stakeholders felt the government should not introduce legislation to intervene in the licensing market. Creative sector stakeholders noted that the government should instead focus on ensuring that the market conditions enable licensing to flourish, particularly by introducing transparency requirements on AI developers in the expectation that greater transparency would enable right holders to better license and enforce their rights.
39. Some expressed concern about who would benefit from licensing and argued for fairer licensing outcomes between large organisations, individuals and SMEs.
40. The market for licensing copyright works for use with AI technology is still new and evolving. We propose not to intervene in the licensing market at this stage. Instead, we propose to monitor the market as it develops and will keep market-led approaches to licensing under review. With regard to AI systems developed outside the UK, we propose to continue monitoring global developments and judicial outcomes.
41. We also propose to identify and assess further levers to support access to valuable datasets, including through the Creative Content Exchange (see Section G of this Report).

Enforcement

42. Effective enforcement is essential to ensure rights are meaningful. The UK is internationally recognised for its strong framework for enforcing intellectual property (IP) rights, but AI may pose new challenges for enforcement.
43. Many stakeholders said transparency obligations on AI developers are a prerequisite for effective enforcement; while some argued that UK copyright law should be applied to models trained overseas.
44. We believe enforcement should be effective, accessible and proportionate. Our approach is to ensure that the UK continues to have a competitive enforcement framework. If any new enforcement measures are considered, they must be accessible to right holders of all sizes, providing effective redress while remaining proportionate to ensure that the UK is able to access the latest technology and innovation.
45. We propose to continue working with partners, including law enforcement and the judiciary, to help ensure the UK enforcement framework remains fit for purpose. We propose further work to identify and address enforcement barriers and consider where action to mitigate these barriers may be required. We propose to consider the case for and approach to regulatory oversight of transparency or other measures, if legislation across these is introduced

Computer-generated works

46. In the UK, there are existing copyright protections for computer generated works created without a human author. This protection, which has been in place since 1988, departs from the core rationale for copyright, which is to encourage and reward human creativity.
47. Most people who responded to this consultation question considered that works created solely by AI should not be protected and supported the removal of this protection, while retaining protection for AI-assisted works. We agree that copyright should incentivise and protect human creativity. We propose to continue to monitor the use and impact of protection for wholly computer-generated works. However, in the absence of evidence of its ongoing value, we propose that this specific type of protection should be removed, while copyright should continue to protect works created with AI assistance.

Digital replicas

48. AI makes it easier to create 'digital replicas' of someone's voice or face. This can be a powerful tool, including for the creative industries. However, when someone's likeness is replicated without their permission it can cause harm. There are some legal protections in place today, but these do not cover all situations where a digital replica is made without consent. There is some support across sectors for enhanced protections for a person's image and voice.
49. However, there was no single view from consultation responses on what form this should take or who they should apply to. We agree that the growing use of realistic impersonation through AI creates new risks for artists and the general public. We propose to explore a range of options for addressing these risks, while protecting the potential of this technology to support legitimate innovation. This exploration will include consideration of whether it would be beneficial to introduce a new digital replica or personality right.

Section B: Copyright and AI systems

1. Copyright enables creators of all types, from writers and artists, filmmakers and coders, database creators and performers, to be paid for their work by controlling how it can be used. It underpins investment in our world-class creative industries, enabling the public to enjoy art, entertainment and information across a wide range of media.
2. From the printing press to the internet, copyright has been shaped by new technology. As media technology has developed, copyright law has grown to embrace it, protecting new types of work and granting new rights to creators.
3. Copyright law seeks to balance protection and investment in new creative content with the need to disseminate information to the public. The way this balance has been achieved has changed as technology has developed. The latest challenge to this balance comes from the rapid rise of generative AI.
4. At the heart of a generative AI model is the “neural network” – a type of machine-learning algorithm. This comprises a number of nodes, arranged in layers from input to output, with connections between them – analogous to biological neurons. By exposing the input layer to billions of examples of creative works – such as text – a neural network can be trained to learn language, concepts and ideas, enabling it to create outputs informed by the content it was trained on, from short sentences, to complete books.
5. These models are frequently trained on billions of copyright works. Often models will be trained in jurisdictions with the most flexible copyright laws. There are many jurisdictions – the USA, the EU, and Japan, where it is lawful (or argued to be lawful), under certain conditions, to train on publicly available copyright works without negotiating a licence with the right holder. We should expect this to be a factor that encourages AI model development inside these jurisdictions relative to those with more stringent copyright regimes, although other factors will be relevant. This, and the fact that their outputs can both be used by and compete with human creators, has presented the copyright framework with new challenges.
6. In this section we consider the process of AI development and the effect of copyright on the access to, and use of, data by developers of AI systems. We also consider stakeholder views on the current copyright framework. The next section considers the potential changes to copyright law that the consultation proposed and sets out the government’s approach.

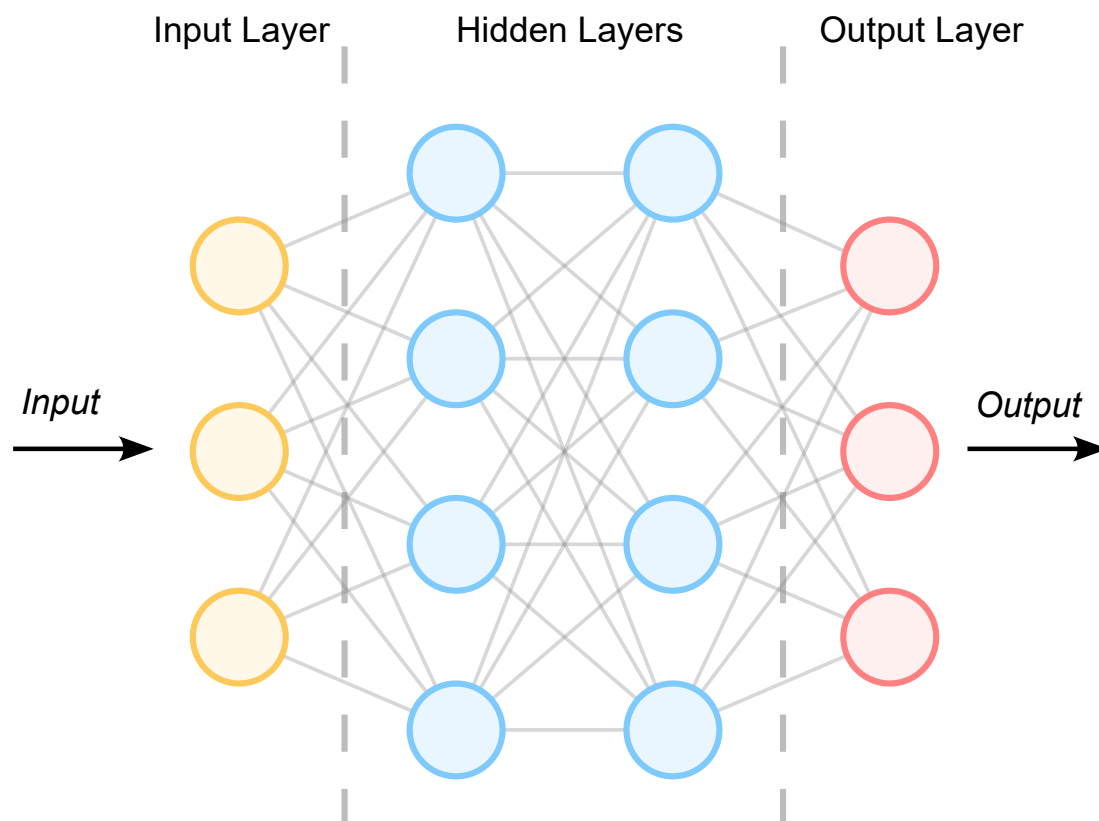


Figure 1: A simplified representation of a neural network

AI system development

AI system design

7. The first stage in AI model development is deciding what problem the model is intended to solve and what outcomes are to be achieved. This guides decisions over the design and architecture of the model.
8. In this section we outline development of a foundation model – a general-purpose model trained on large datasets which can be fine-tuned for many different applications. This is a simplified description and there are many other types of AI model and development process. Our aim is to illuminate the types of data processing that may take place during development. Subsequently, we consider how copyright is relevant to each of these stages.

Data acquisition

9. Foundation models are trained using very large datasets – often including billions of items of data. This data may be structured (for example, labelled) or unstructured. Data may be obtained from a range of sources but given the need for such large volumes it is often obtained through the use of text, images, and other content which has been made available on the internet, as well as being supplemented by other large datasets.
10. Various different types of dataset are explored below, where we consider how these datasets are assembled and licensed.

Pre-processing

11. The inputs to a model will reflect the type of output that the model is intended to produce. A large-language model (LLM) – a model capable of understanding and generating text – will be trained on billions of items of text, such as webpages, and computer code. An image model will be trained on billions of images – such as photographs and drawings.
12. This input data needs to be pre-processed before it can be used for training. This may involve cleansing, standardising, tokenising, and compressing the dataset. Data cleansing may include various steps such as removal of repeated copies (de-duplication), removal of content which is harmful or explicit, and filtering low-quality content or poorly described labels. It may also include enforcing consistency in size or format of content and potentially reducing the dataset to focus on specific types of content. Data cleansing may be performed by the developer, but many datasets will already have undergone some cleansing by their providers.
13. Before use in a model, data will be processed so it can be more easily digested. This may include tokenisation and compression. Tokenisation is where input data is broken into smaller standardised units, such as words or sub-words, then represented as numbers which can be understood by the model.
14. Content may also be compressed for easier processing. One approach uses a neural network called an autoencoder, which translates input data into “latent space”. This breaks down input data into key features, such as shapes and textures. An image of a house may include features like bricks, which are repeated in the image and are common to other images of houses. The network will identify these common features and store them, rather than the individual pixels in the image. This allows for a high degree of compression. It is often these compressed representations of images, rather than the images themselves, which are used for later training of an AI model.



Figure 2: Text is tokenised – broken into chunks which are assigned values – in preparation for AI model training

Training

15. AI model training is an iterative process, which is typically broken into stages. Each stage refines the model further, to improve its ability to accomplish a desired task. Each stage may require more data to be input to the model, from different sources, and by different parties.
16. Below, we split training into two halves – “pre-training” and “fine-tuning”. This echoes terminology frequently used by foundation model developers. But it is a simplification of the process, which will often involve different stages, depending on the type of model under development.

Pre-training

17. Training of foundation models will typically begin with a “pre-training” stage. Pre-training is a crucial stage in the development of foundation models.
18. This stage requires vast quantities of data and massive computational resources. One reason is that it involves self-supervised training. Instead of a person labelling data, the model makes sense of the data it is exposed to by itself.
19. This is typically done by giving the model prediction tasks. A transformer model may be given the start of a sentence and will be asked to predict the next word, or it may be given a sentence with missing words and be asked to fill them in. The model will work out a suitable word based on the large number of examples it has been trained on. An attention mechanism enables it to distinguish between similar phrases used in different contexts.⁴
20. Another type of model is a diffusion model, which is often used for image generation. When training a diffusion model, noise is progressively added to images, over a series of steps. The model is presented with these images and asked to predict the noise in each, so it can reverse the process. Its goal is to produce a denoised image which resembles the original. At each step, the parameters of the model are adjusted according to how close the denoised image is to the original image. Once trained, a model is able to completely reverse the noising process, producing an image from pure noise.
21. Images may be paired with text descriptions during training, allowing the model to learn how words relate to visual features. Such a model is then able to generate an image from scratch in response to a text prompt.

Fine-tuning

22. Pre-training involves large amounts of data from a broad range of sources and is often unsupervised or semi-supervised. To ensure a model can give high-quality results in response to a specific task, and meets other conditions (including regulatory requirements), it will be followed by one or more fine-tuning stages. This training will focus on domain-specific data – data that is relevant to the specific task that the model will be put to.
23. For example, ChatGPT is fine-tuned from the GPT model, to enable it to fulfil its task as an AI assistant. It is trained through a supervised process, where a labeller will first train a policy (a set of rules for responding prompts) by selecting a prompt and demonstrating to the model what an output for that prompt should look like. Then they rank outputs from best to worst. This information is used to train a reward model, which can further optimise the policy by ranking outputs by itself.
24. Fine-tuning may be based on a new dataset of new content, acquired from an external source. Or it may be based on a subset of the initial dataset or be generated by the developer themselves. Any new content is also likely to need pre-processing before it is used for fine-tuning.

⁴ OpenAI’s GPT model is an example of a model trained in this way – GPT stands for “General Pre-Trained Transformer”.

Release

25. When a model is performing to the required standard, the model – or services based on it – will be released to the public.
26. Some models are provided to the public on an open-source or open-weight basis. Open-source models generally allow developers to view the weights, code and architecture of a model, and to adapt and build on them. Open-weight models allow access to weights but have more restrictive conditions on how they can be used. A popular platform for such models is HuggingFace, which hosts a wide range of models and datasets.
27. Other models are provided on a closed-source basis, with application programming interfaces (APIs) allowing interaction with the model. These allow customisation and fine tuning, and permit integration into different services. But they do not provide as much flexibility as open-source models, and the inner workings of the model are not visible.
28. Each of these approaches allows a third party to take a trained model and make a service which is based on it.
29. As well as providing a model to the public for further development, the model developer may provide their own services directly to the public. Examples are Google, OpenAI, and Meta's AI assistants. These are based on their own AI models, which have been fine-tuned and integrated into a service before release.

Task-specific training and development

30. A remarkable quality of general-purpose foundation models is the wide range of tasks that they can be applied to. A model which has been made available to the public (whether open or closed source) will be further refined and deployed by a range of parties for specific applications.
31. Many developers in the UK will take a base model developed in another country (such as the USA or China) and further fine-tune it, or add additional components to it, which allow it to accomplish a specific task.

Post-release fine tuning

32. At this stage, a developer may fine-tune a pre-trained model on a further, more specialised dataset. For example, it may be trained on a corpus of legislation and case law to support legal research applications. Fine-tuning enables the model to consistently deliver the specific output that is required by the service.
33. This is often the type of AI training and development which takes place in the UK. It often engages micro-businesses and startups, or researchers in research organisations and spin outs. In its consultation response, The Startup Coalition indicated that many UK startups operate at the "application layer of the stack", with many of these companies fine-tuning existing models for task-specific applications.
34. Depending on its objective, fine-tuning may rely on use of a further dataset with a large number of datapoints – for example, a specialist medical dataset for a diagnostic application.

35. There are ways to reduce the time and volume of data needed to fine-tune a model. For example, Low-Rank Adaptation (LoRA) works by adjusting a smaller matrix and adding to existing model weights. This is faster, cheaper, and can often be achieved using much smaller datasets than standard model training.

Retrieval-Augmented Generation (RAG)

36. This is an approach where a pre-trained model is connected to an external knowledge base to ensure outputs are relevant, accurate and up to date. It is commonly used for AI search and by AI assistants. RAG can help mitigate the risk of hallucinations by a model and improve its factual accuracy.
37. For example, a company may use an AI assistant to help staff understand its policies. To give the right results, the AI assistant will search the company's records, and generate a response based on what it finds.
38. Many online search engines now provide AI search assistants, which use RAG to provide search summaries, using web indexes as their data source.

Deployment

39. AI models may be deployed in a wide range of services for end-users. When a service outputs content of a similar nature to creative content it was trained on it is often referred to as "generative" AI. For example:
- An image-generation service, which generates realistic images from text prompts;
 - A song-generation service, which generates music and vocals.
40. Some services may allow or require further content to be input at this stage. For example:
- a photo editor, which allows a user to upload their photos for editing;
 - an AI assistant which allows a user to upload a document for it to summarise.
41. Other services trained using text, images and other content will not generate similar content, but will output other information or support other goals. For example:
- an app which identifies plants and fungi and outputs information about them
 - a content moderation service, used to remove undesirable content from a social media platform
 - a breast cancer screening service, which is able to detect and flag potential anomalies for review.
42. An essential component of many of these stages, from training to deployment, is a large quantity of data, in the form of text, images, and other content. This content is likely to be protected by copyright and subject to copyright law.

Copyright law

43. Copyright protects works such as original literary, dramatic, musical and artistic works, including illustration, photography, songs, plays, software, web content and databases. Copyright also protects sound and music recordings, film and broadcasts. By providing exclusive rights to the creators and producers of these works, copyright allows them to control who uses their work, and under what conditions. This enables creators to be rewarded for their work and encourages investment in new content.
44. Copyright has been described as the lifeblood of the creative industries, without which they would be unable to thrive and grow. But it also applies more widely, covering any work with sufficient creative expression. This will include more functional works such as academic papers, computer software, and government reports, as well as informal works such as personal letters and photographs.
45. The history of copyright is one of adaptation to new technologies. The first copyright act – the Statute of Anne, enacted in 1710 – applied only to literary works, granting authors rights in printed books. Subsequent acts applied to engravings, sculptures and paintings. In more modern times, copyright has further adapted to cover media such as films, broadcasts and sound recordings, and to respond to the challenges of digital copying and the internet.

Types of protection

46. The modern copyright law of the United Kingdom is set out in the Copyright, Designs and Patents Act 1988 (the “CDPA”). There are two types of copyright in UK law – “authorial” and “entrepreneurial”. Authorial copyright applies to original literary, artistic, musical, and dramatic works. It protects an author’s⁵ creativity, as expressed in their work. Authorial copyright protects the work in any form. A novel will be protected by copyright as an original literary work, but it will also be protected if it is adapted for a film, television programme, videogame, etc.
47. Authorial copyright aims to encourage and reward creativity. As such, it protects an author’s creative expression in a work, but not factual information, methods, or concepts. A news article will be protected by copyright because a journalist will have made creative decisions while writing it, which are expressed on the page in their choice of words. But the facts that they are reporting can be repeated by another journalist, in another article, without infringing copyright, provided the creative expression in the work is not copied.
48. This distinction between protected creative expression and unprotected ideas, concepts, and facts – sometimes called the “idea-expression dichotomy” – is found in copyright laws around the world.
49. Entrepreneurial copyright (in international law referred to as “related” or “neighbouring” rights) protects individual embodiments of certain types of content – sound recordings, films, broadcasts, and published editions. These rights protect a

⁵ In UK copyright law, creators and producers of copyright works are referred to as “authors”, regardless of the nature of the work.

specific recording, broadcast, or typographical arrangement, regardless of its originality. However, they only protect the specific embodiment of sound, image or text. The owner of sound recording rights in a specific recording of a song cannot use those rights to control other recordings of the same song.

50. As well as copyright, similar rights exist in performances which are captured in films or sound recordings, and in databases in which there has been a significant investment.
51. A single item of creative content may be protected by multiple rights belonging to different people. For example, a pop song will comprise a musical work (composition) and literary work (lyrics), each of which may be authored or owned by different people. The song can be recorded by different artists, with the producer of each sound recording, and the performers on it, benefiting from separate protection. Right holders can be individuals, but also businesses of different sizes.
52. In this report, we often use “copyright” as a shorthand for all of these different types of protection, given their similarities. Where the differences between them are relevant, we aim to highlight them.

Rights and exceptions

53. Copyright law gives the author of a work, or a person to whom they have transferred or licensed their rights (such as a producer or publisher), the right to authorise or prohibit certain uses of that work. These restricted acts, set out in section 16 CDPA, include rights to copy the work, to distribute copies to the public, to perform a work in public, and to communicate the work to the public – including making it available on the internet.
54. Copyright law contains a number of exceptions, which aim to support wider public policy goals. These are part of the balancing act that copyright law aims to achieve between protecting and promoting creativity and allowing access to and use of information. For example, there are exceptions for quotation, news reporting, use in education, use by libraries. Relevant exceptions to AI training include the exception for temporary copying (section 28A CDPA) and the exception for data mining for non-commercial research (section 29A CDPA).
55. Although exceptions will apply to some specific uses and applications, in general permission will be required from a right holder before a work can be used for a restricted act (section 16 CDPA). Permission is commonly granted through a licence agreed between the right holder and the user. Licensing is described in more detail in Section G of this report. Enforcement is covered in Section H of this report.
56. Copyright law confers moral rights as well as economic rights. These include the rights of authors and performers to be identified as such and to object to derogatory treatment of their works or performances. Moral rights cannot be assigned by an author or performer, though they may be waived.

Copyright in other countries

57. Many modern technologies and digital content services are developed across multiple territories. Copyright is territorial, which means that, unless agreed otherwise, the copyright law of the country where a relevant act takes place will apply. That means that UK right holders whose works are accessible in other countries will find that use of those works in those countries is subject to local copyright laws. The development of AI models and systems outside the UK, under foreign copyright laws, can create challenges for UK right holders.
58. There are several international treaties on copyright and related rights. These set out common minimum standards that their parties should adhere to, including who they must give rights to, for which type of content, for what uses, and for how long. Most countries belong to some or all of these treaties, including leading AI economies such as the USA, China, and the EU and its member states.⁶
59. These treaties lead to a high degree of convergence between national laws on the types of work protected and the rights available under copyright and related rights. This in turn helps to support international trade and investment. But the treaties also allow for some flexibility in national implementation.
60. One such flexibility is the ability to provide for exceptions to rights. However, any exceptions must comply with the “three-step test”. This test, which is found in several international treaties, permits members to provide exceptions to copyright and related rights as long as they:
 - i. are confined to certain special cases;
 - ii. do not conflict with a normal exploitation of the work; and
 - iii. do not unreasonably prejudice the legitimate interests of the right holder.
61. Several countries provide exceptions in their copyright law which are relevant to AI training and development. This includes the EU, Japan, and Singapore, which all have specific “text and data mining” (TDM) exceptions. Some other countries, notably the USA, allow uses relevant to AI training under a “fair use” principle, though the scope of fair use is contested. Other countries have announced different approaches, or are developing their views, including Australia, Canada, India, and South Korea.
62. The EU provides two exceptions for TDM⁷ which have been interpreted to include data extraction for AI training and fine tuning. A first exception covers TDM for scientific research, and a second is a general-purpose TDM exception with a right holder opt-out. For online works, this opt-out must be exercised in machine-readable form. Both exceptions require “lawful access” to the works which are used. Judgments interpreting aspects of these exceptions have been published in a

⁶ These countries are all party to at least the TRIPS agreement, the Berne Convention, WIPO Copyright Treaty, and WIPO Performers and Phonograms Treaty.

⁷ Articles 3 and 4, EU Directive 2019/790 on copyright and related rights in the Digital Single Market <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>

number of member states, and a preliminary reference has been made to the Court of Justice of the EU, but a number of aspects remain unclear.

63. Japan provides an exception for TDM,⁸ which can apply to AI development. It allows a work to be used provided that it is not intended to enjoy or cause another person to enjoy the “thoughts or sentiments expressed in the work”, and the use does not unreasonably prejudice the interests of right holders. Guidance issued by the Japanese Agency for Cultural Affairs⁹ indicates that it would not apply to generation of material “similar” to protected works, including RAG. Nor would it apply to use of material which is provided in a format expressly for information analysis, such as a commercially supplied dataset. Finally, it would not cover the use of material which is restricted using technical measures including solutions such as Robots Exclusion Protocol.
64. The Singapore exception¹⁰ allows for computational data analysis, such as sentiment analysis, TDM, and training machine learning systems. The user must have lawful access, but this allows use of infringing copies of works, if the user was not aware of this status. There is no distinction between commercial and non-commercial use.
65. India does not currently have a specific TDM exception. India’s Department for Promotion of Industry and Internal Trade published a working paper in December 2025 proposing statutory licensing, which acts as a mandatory blanket licence, and in effect provides a broad TDM exception for AI training on all lawfully accessed copyright content.¹¹ AI developers would be obliged to pay for this licence, with a government agency responsible for collecting payments and distributing them to Collective Management Organisations (CMOs), which would in turn distribute them to right holders. A consultation ran until 6 February 2026, and it is yet to be seen how the Indian government will respond.
66. Canada does not have any specific TDM exceptions but is currently considering options following a consultation it ran between October 2023 and January 2024. Australia announced in October 2025 that it would not introduce a TDM exception but will continue to work with a range of stakeholders from the creative industries and AI sectors in areas of joint interest.
67. The USA, South Korea, and Israel provide a “fair use” defence, as do some other territories. Under this principle, works may be used without permission if the use is considered “fair”. The USA leads the fair use territories in developing the principle in the face of new technological challenges.¹²

8 Article 30-4 Japanese Copyright Act <https://www.cric.or.jp/english/clj/cl2.html>

9 General Understanding on AI and Copyright in Japan https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf

10 Section 244 Copyright Act 2021 <https://sso.agc.gov.sg/Act/CA2021?ProvIds=P15--pr244->

11 One Nation One License One Payment: Balancing AI Innovation and Copyright (December 2025) Working Paper on Generative AI and Copyright, Department for Promotion of Industry and Internal Trade, Ministry of Commerce and Industry, Government of India <https://www.dpiit.gov.in/static/uploads/2025/12/ff266bbeed10c48e3479c941484f3525.pdf>

12 U.S. Copyright Office Fair Use Index <https://www.copyright.gov/fair-use/>

- 68.** Many AI models have been trained in the USA by developers who cite fair use as a defence to their use of copyright works in training. Fair use is a defence which is assessed on a case-by-case basis by the courts in light of four factors:¹³
- i.** the purpose and character of the use (including whether it is commercial or non-profit). This includes whether the use is “transformative”;
 - ii.** the nature of the copyright work, for example whether it is highly creative or functional;
 - iii.** the amount and substantiality of the portion of the work used in relation to the work as a whole; and
 - iv.** the effect of the use upon the potential market for or value of the copyright work. This includes whether the use would substitute for, or damage the market of, the original.
- 69.** Fair use questions relevant to AI remain unsettled. Over 70 cases involving copyright have reportedly commenced before the USA courts, most involving fair use. It is beyond the scope of this report to consider fair use, and the extent to which it may apply to AI development, in detail. A description of fair use and how its factors might be applied to AI development can be found in a 2025 report on Generative AI Training from the U.S. Copyright Office.¹⁴ As will be seen in the rest of this report, the outcome of these cases will have consequences not only in the USA, but in the UK and around the world.
- 70.** In summary, countries’ copyright laws take different approaches to the use of copyright works to train AI systems. Several major economies are considering the introduction of new laws and in others the effect of existing laws is disputed. The lack of a broad data mining exception or fair use principle in UK law places the UK among the countries with greater protections for right holders, and less flexibility for AI developers.

Copyright and artificial intelligence models

- 71.** Above, we considered the development of an artificial intelligence model from its creation to use in a variety of applications. Copyright works may be used by a model at various stages in its life.
- 72.** In this part of the report, we consider how copyright applies to AI models, and to the data used by them, at these different stages. Many types of AI system exist. For our illustration we continue to use the case outlined above: a general-purpose foundation model which is fine-tuned for specific applications. This illustrative example has been chosen as it demonstrates a range of uses of copyright works. In practice different steps may be taken depending on the model type and use case.

¹³ Section 107 Copyright Act 1976 Chapter 1 - Circular 92 | U.S. Copyright Office

¹⁴ U.S. Copyright Office report on Generative AI training <https://www.copyright.gov/ai/>

- 73. We consider the effects that copyright has on the access to, and use of, data by AI developers, and the opportunities for control and remuneration it provides to creators. We also summarise relevant responses to the consultation.
- 74. This is the “status quo” option (option 0) in the government’s consultation. The description below is informed by responses to the consultation, as well as other sources, and is followed by a summary of those responses. Potential economic impacts on right holders, AI developers and AI users under the status quo are considered further in the accompanying impact assessment.

AI system design

- 75. At the heart of an AI system is a statistical model which is trained using algorithms. Neither pure statistical models nor algorithms are protected by copyright, which does not extend to ideas, methods, or mathematical concepts as such. However, the implementation of an AI model, and the broader AI system, will be written in computer code. Computer code can be protected by copyright as a literary work (section 3(1)(b) CDPA), to the extent that it is a creative expression of its coders. However, copyright does not protect its functionality.

Data acquisition

- 76. A general-purpose foundation model will require billions of data points for its initial training. The data and information required will inevitably include copyright works, which may be obtained from a range of different sources. Each type of source has different copyright implications.

Licensed data sets

- 77. Many licensed datasets for AI training are now available. The content in such datasets may be owned by one or multiple right holders, and it may be provided under a range of licensing terms.
- 78. A licence will provide the necessary rights for data processing and training including to do acts normally restricted by copyright. This will include the right to make reproductions of relevant works under section 17 CDPA. It will also typically place conditions on the use of material – for example it may require the uses of trained models to be limited to certain purposes. It may also impose attribution, reporting, or auditing requirements.
- 79. For example, in its consultation response, Getty Images said, “we enter bespoke license agreements with carefully defined AI usage rights and with appropriate additional safeguards included, ensuring that all licensed content is allocated a share of licensing fees.”
- 80. The underlying rights in these datasets will often have been obtained, or licensed, from original creators and performers, sometimes through specific contracts for that purpose, and sometimes through broader agreements where rights are licensed in bulk to be aggregated by a third party. Examples of rights aggregation include collective licensing and image libraries. Commercial licensing and developments in the AI licensing market are described in detail in Section G.

81. There are also non-commercial, freely licensed datasets, such as those hosted by Wikimedia. These may be licensed under open licences, such as Creative Commons licences. While such licences may permit free use of datasets, they often have some downstream requirements. In its consultation response, Wikimedia noted that its materials could be used for machine learning, but that attribution was a core principle of its licences.
82. Another source of licensed data is through end-user agreements with users of services. This includes social media platforms and message boards. For example, Reddit has announced partnerships with Google and OpenAI that allow them to train their AI on Reddit users' posts, which is also outlined in its terms of service.¹⁵ By signing up to such services, a user will grant the platform a licence to use their posts for various purposes, which may include the right to train AI models. Platforms may provide users with ways to opt out of AI training on their posts, but it is not always straightforward.

Unlicensed data sets

83. Many foundation models are trained on large datasets of images, text, and other content, obtained from the internet via methods such as web-crawling or crowdsourcing. While some of this content will be out of copyright or made available for training under licence (for example, via an open licence), much of it will not have been licensed for this purpose. This practice is at the heart of current legal disputes and policy debates.
84. Often, the content is curated and assembled by a third party, who makes it available to AI developers. This may be done directly, by storing and hosting the content (for example, the Common Crawl, which scrapes billions of websites, and hosts these pages), or indirectly, by linking to content uploaded elsewhere (for example, the LAION datasets, which contain image-text pairs, but link to and do not host the actual images). Some datasets contain only raw data, whereas others are curated, with volunteers labelling content (for example, ImageNet).
85. There are also combined datasets, such as the Pile, which combines multiple online datasets into a single, diverse set. This helps meet the objective of training LLMs on a large amount of diverse data. The Pile included unlicensed material such as Books3 – a dataset of pirated books. Books3 has since been taken down, but its use (through the Pile) has been cited in several legal cases in the USA.¹⁶ Other combined datasets, such as the Common Pile, claim to be fully openly licensed.¹⁷
86. If a dataset is assembled in the UK, it is likely to involve reproductions under section 17 CDPA, which will infringe copyright unless an exception applies. As noted above, exceptions in the UK only apply in specific and limited cases. But often the curation and assembly of content will have been done outside the UK, and those assembling them may seek to rely on local exceptions.
87. Regardless of the law under which a dataset is assembled, if a person makes it available to the public in the UK, this may constitute an infringement by

15 <https://www.businessinsider.com/reddit-openai-deal-ai-data-partnership-2024-5>

16 <https://uk.pcmag.com/ai/153978/a-pile-of-pirated-work-authors-sue-anthropic-ai-joining-youtuber-backlash>

17 <https://arxiv.org/abs/2506.05209>

communication to the public under section 20 CDPA. A key question is, where links are provided to works rather than the works themselves, whether doing so makes those works available to a “new public” who would not otherwise have access to them.

Data download

- 88.** Once a dataset of creative works is acquired, local copies of those works will be made by the person using them for AI training. If downloaded in the UK, section 17 CDPA (infringement of copyright by copying) will be relevant. This section applies to every type of copyright work and to copies stored by any means. Unless licensed or covered by a specific exception the person making copies will infringe copyright.

Effects on creators and AI developers

- 89.** As described above, many of the materials that form part of datasets are protected by copyright, and many of the acts involved in their assembly and provision may infringe copyright if done without a licence in the UK. Copyright law enables datasets to be licensed, so that right holders are paid. But dataset providers and users of datasets may also consider their own commercial and legal interests and choose to conduct their activity outside the UK to take advantage of exceptions or other ways to reduce their liability. It may be easier for large AI developers to take advantage of other countries’ laws. SMEs or individual developers may find it difficult to take similar advantage, given the resources needed to research and ensure legal compliance.

Pre-processing

- 90.** Further copies of protected works may be made during pre-processing – when a dataset is put into the right format for training. This may involve large scale temporary copying of the works in the data set.
- 91.** Pre-processing may include compression. In copyright, a copy is a copy whatever its form and whatever medium it is stored in (section 17(2) CDPA), which may include compressed copies. The crucial question is whether the original creative expression in a work is retained.
- 92.** Take a drawing of a cartoon mouse. Many of the broad features will be generic – it may have a tail, ears, whiskers, etc. But creativity will be expressed in the shape of these features, and how they are combined together. An image of the drawing may be highly compressed, but as long as it still includes these expressive elements it will still constitute a copy of the work. It is a matter of debate whether latent space and other lower-order representations of content used to train AI models contain such expression.

Training

Pre-training

- 93.** As noted above, pre-training can be done in a variety of ways, but each entails adjustment of model weights, based on feedback, to produce a desired result. Pre-training is typically self-supervised rather than manually labelled.

94. There are various points in this process at which a copy of a work may be made, which may infringe copyright unless licensed or covered by an exception. Some developers have cited the temporary copies exception (section 28A CDPA) and the non-commercial research data mining exception (section 29A CDPA), though these contain a number of conditions and will only be applicable when those conditions are met.
95. Individual images, text or other inputs (or compressed versions of them) are not stored in the model like they are stored in a database. Instead, weights are adjusted each time an image is presented to the model. If multiple datapoints (images, text extracts etc.) contain similar content it will reinforce certain weights.
96. Each individual image will have a limited influence over the weights in the model. As a consequence, a model is often said to learn patterns and relationships, rather than storing data or content as such. In its consultation response, Google described generative AI models as “not databases or information retrieval systems” but networks storing “learned relationships” between input tokens.
97. According to this viewpoint, which is contested by other stakeholders, the patterns and relationships stored in the model are outside the scope of copyright protection. Copies of works might have been made during the training process but are not necessarily “stored” following training.

Memorisation

98. This is not to say it is impossible for a model to memorise a work. A reason for memorisation is “overfitting” in the training process. Overfitting is when a model is over-exposed to certain types of input, so finds it difficult to generalise to examples that it hasn’t been exposed to.¹⁸
99. Take the cartoon mouse. A dataset may have been de-duplicated to avoid identical images, but the same mouse may appear in different images. Training on a sufficient number of these will make it more likely that specific features of the mouse are memorised and capable of being accurately reproduced by the model. This risk increases the smaller and more repetitive the dataset.
100. The extent to which memorisation is possible will depend on a model’s design and training process. Model developers often say their intention at the pre-training stage is a model which is able to generalise well, which means avoiding overfitting and memorisation.
101. In *GEMA vs Open AI*, the Munich Regional Court held that song lyrics that had been output by an AI model were evidence that those works had been memorised by it. In its submission to the consultation, the Publishers Association pointed to evidence of a model reproducing verbatim text from a Harry Potter novel. The presence of a work, or an extract of one, in a model’s output suggests that it may have been memorised by the model because of repeated exposure to it.

¹⁸ A description of the relationship between memorisation and overfitting is set out in the agreed expert evidence cited at paragraph 144 of the *Getty Images vs Stability AI* judgment.

Location of pre-training

- 102.** In their consultation responses, large AI firms indicated that the training of their foundation models takes place outside the UK – usually in the United States. Copyright is territorial, meaning the applicable law is that where a relevant act took place. A person training a model will need to comply with the copyright law of the country where copies are made, which may influence their choice of training territory.
- 103.** Often companies training models in the USA will cite the USA fair use principle as a defence to copyright infringement. A considerable number of fair use cases are currently before the USA courts – the U.S. Copyright Office referred to “dozens of lawsuits” in its report on Copyright and Artificial Intelligence published in May 2025.¹⁹
- 104.** Other factors relevant to the choice of training territory include the location of computing resources, energy costs, support for startups and SMEs, and the wider regulatory environment.

Fine-tuning

- 105.** The object of the fine-tuning stage is to take the generalised model and train it on domain-specific data, in order to achieve certain task-specific outcomes. For example, the GPT model is fine-tuned by OpenAI to produce ChatGPT.
- 106.** For a generative model, the new dataset may comprise content such as images, music, or text, which is likely to be protected by copyright. A supervised learning process may be employed, with labelled examples and human feedback, in order to direct the model towards task-specific outcomes.
- 107.** The same copyright questions arise in relation to this new dataset as for the original, more general dataset. The new dataset will also have to be acquired and pre-processed, which may engage copyright as described above. Copies of works are likely to be made during the fine-tuning process and, unless a relevant exception applies, these will need to be licensed.
- 108.** Because of its task-specific nature, a fine-tuning dataset is likely to be more specialist, more homogenous, and (relatively speaking) smaller than the original training dataset. As such, it may be less likely to be available over the internet, which may support licensing by right holders. Examples of specialist datasets that have been licensed by foundation model providers include up-to-date information from news publishers.
- 109.** In our example, we assume the entity pre-training and fine-tuning a foundation model is the same. In which case, they are likely to be located in the same jurisdiction, and that country’s law will be applicable. As noted above, the applicable law will often be that of the USA or another country, rather than the UK.

19 U.S. Copyright Office report on Copyright and Artificial Intelligence, Part 3: Generative AI Training pre-publication version <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>

Release

- 110.** As noted above, although works may be copied during training, it does not necessarily follow that a particular work will be memorised by an AI model or otherwise stored in an AI system. However, if copies are stored in the system, and those are infringing copies, then a person who deals with that system in the UK may be at risk of copyright infringement. For example, if a person knows or has reason to believe a system contains infringing copies then they can be liable for importing, possessing or dealing with those copies (sections 22 and 23 CDPA), or other acts of secondary infringement.
- 111.** An article will be considered an infringing copy for the purposes of the secondary infringement provisions if making it constituted infringement of copyright in the UK, or if it was made somewhere else but in circumstances which would have constituted a copyright infringement if it had been made in the UK and the article has been or is proposed to be imported into the UK (section 27 CDPA).
- 112.** In *Getty Images vs Stability AI*²⁰, it was held that a trained model could, in principle, qualify as an “article” and so could be an infringing copy (even if it had been trained abroad). However, for that to be the case, the trained model would need to contain or reproduce a copy of a relevant work. As there was no evidence in that case that the specific model (Stable Diffusion) contained any relevant works, it was held not to be an infringing copy. As such, the secondary infringement provisions did not apply. But the outcome may be different in other cases, if there is evidence that a model does contain copies of a work. This is the first time a court has ruled in relation to copyright and AI training in the UK, and its interpretation of the law in this area is currently subject to appeal.
- 113.** As the judgment illustrates, one challenge facing right holders who seek to make secondary infringement claims will be demonstrating that a copy of their work is present in, or reproduced by, a model.

Task-specific training and development

Post-release fine-tuning

- 114.** As noted in the previous section, foundation models are often used by a downstream developer as the basis for a task-focused model or service. This will often involve further fine-tuning of the model, which may include training on a new dataset.
- 115.** If a developer needs to make copies of the new dataset during fine-tuning, then the same copyright considerations will apply as in the initial training stage. Unless a specific exception applies, they will need to license copies made during the training process for example, downloading the dataset, pre-processing it, and introducing it to the model.

²⁰ *Getty Images v Stability AI* [2025] EWHC 2863 (Ch) <https://www.judiciary.uk/wp-content/uploads/2025/11/Getty-Images-v-Stability-AI.pdf>

- 116.** Again, as above, once trained the model may not store copies of the works it was fine-tuned on as such, depending on the approach taken. What may be different compared to the pre-release model training described above is the specialist nature of the dataset and the location it is used in. UK-based firms fine-tuning on specialist datasets are less likely to be able to conduct this activity in another country, so UK copyright law may become more relevant at this stage.

Retrieval-Augmented Generation (RAG)

- 117.** Another technique used in AI-based services is RAG. This is when a model responds to a user request for information by searching an external database (which may include the internet, company records, research papers, etc.) and incorporating this information into its responses.
- 118.** Assuming the external database comprises copyright-protected content, and copies of it are downloaded locally within the UK for processing, section 17 CDPA will apply. Those copies will need to be licensed, unless the use falls within an exception.
- 119.** Certain activities may be considered to fall within an implied licence. For example, if a person sends an email to a company, it is likely to be implied that the company can store copies of that email and use tools to search those copies.

Deployment

- 120.** Copyright may also be relevant when an AI-based service is deployed by an end user. Examples include:
- A service may generate a substantial part of one or more copyright works in its output. If done without a licence, this could infringe the rights of reproduction, communication to the public, or other rights. Depending on the circumstances, both the user and provider of a model may be liable for infringement of copyright. For example, if a user specifically requests an AI service to reproduce an artist's work and the service does so, each may be liable.
 - A service which outputs wholly AI-generated content in response to a prompt may be protected by copyright as a "computer-generated work". Section I of this report covers issues relating to this category of copyright work.
 - A service may use additional works at the point of inference. For example, a copy of a work may be input to photo-editing software with AI-enhanced functions. If the work was not created by the user, they will need to ensure they have the relevant permissions to use it, including to make copies of it.

Status Quo: Views from consultation

- 121.** In December 2024 the government launched a consultation on copyright and artificial intelligence. We received 11,520 responses: 10,110 via Citizen Space and 1,410 by email. A significant proportion of responses either repeated or built-upon a set of answers provided by organisations and individuals in this area. These are all recorded as separate responses.

- 122.** The consultation asked a mixture of closed (multiple-choice) and open text questions. Throughout this report, the percentage of respondents answering each multiple-choice question are based solely on the Citizen Space responses. However, for question 2, where respondents were asked which option they prefer and why, effort has been made to record the preferences of those answering by email. This differs from the figures reported in the Progress Update²¹, which only counted the preferred option of those responding through Citizen Space.
- 123.** The status quo option in the consultation was option 0. It received support from 10% of respondents, representing a diverse range of stakeholders including creative industries, academic institutions, and some technology firms. Many of the respondents supporting option 1 – stronger copyright – also supported the status quo in copyright law. The stronger measures they sought tended to relate to copyright enforcement – notably transparency measures – rather than changes to copyright law itself.
- 124.** Numerous creative industry respondents argued that, as licensing is possible under the current law, there is no need to change it. Several pointed to examples of licensing, for example by image libraries, publishers, and record labels.
- 125.** Some respondents highlighted uncertainty created by proposals for change and said this could potentially harm the nascent licensing market for AI training data. For example, the Society of Authors stated that: “the present options and consultation documents are actually creating uncertainty – and this has a direct adverse impact on the current market for rightsholders.”
- 126.** The consultation did not ask any specific questions about enforcement against unlicensed text and data mining under current legislation. However, many respondents expressed a desire for more stringent enforcement action against AI developers for infringement by training AI models on copyright materials, regardless of any changes to the law. Enforcement is covered in Section H of this report.
- 127.** Respondents who supported option 0 (and supported the status quo copyright law under option 1) generally shared the view that the most important objective of any future changes should be to ensure that AI developers must seek permission and acquire licences for the use of copyright works. Many responses expressed strongly negative opinions towards development of AI in general, and its potential impact on the creative industries. Others were supportive of AI and noted potential benefits to the creative industries but criticised the practices of AI developers.
- 128.** These included views from individual right holders, and SMEs. For example, the Creators’ Rights Alliance, an umbrella organisation representing creators from a range of sectors, noted challenges faced by individual right holders, including writers, illustrators and photographers, who wish to prevent or license the use of their work by developers under the status quo. They noted that it can be difficult to engage with large AI firms, including when approaches are made by representative bodies. They also said that few licensing deals have been reached which deliver benefits to individual creators, and remuneration when agreed will be very small.

21 Copyright and AI progress report, December 2025: <https://www.gov.uk/government/publications/copyright-and-artificial-intelligence-progress-report>

- 129.** SME right holders also noted challenges under the status quo. For example, the Association of Independent Music supported licensing under the status quo but noted that (at the time of writing) only “very few smaller AI developers” had approached independent record labels to do so. Other SMEs, such as small publishers, also noted difficulties in engaging with AI companies to secure licences and enforce rights under the status quo. But there were also positive comments about licensing by SMEs under the status quo. For example, the Independent Publishers’ Guild, noting that 98% of the publishing sector is SMEs, said that 39% of their members had already licensed content to AI companies or were in active discussions to do so. Further discussion of licensing and its effects on different groups is contained in the licensing section of this report.
- 130.** AI developers and users also noted challenges under the status quo, which were said to have particular impacts on SMEs and individuals. As noted above, many developers and users of AI in the UK will be SMEs and individuals, which build on and apply systems developed by larger firms (often outside the UK).
- 131.** For example, Tech UK, representing the UK technology sector, noted that partnerships between AI developers and content holders have been growing, but that their smaller members noted challenges accessing data at scale under the status quo. They noted that interventions which improve access would benefit smaller (SME and individual) developers in particular. They raised the benefits for innovation of a diverse ecosystem of different-sized businesses in which smaller developers were able to participate. Users of AI systems also noted challenges under the status quo. For example, the BioIndustry Association, representing life sciences businesses in the UK, noted current challenges over accessing datasets for use with AI systems, including uncertainty about terms of use, which may have particular impact on the 75% of their members which are SMEs. Other user groups, including in the research and cultural heritage sectors, noted similar challenges.

Status Quo: Assessment

- 132.** We note the endorsement by creators and right holders for the UK’s current copyright law under both option 0 and option 1. In the absence of any new legislation, the status quo will continue, and we will continue to monitor its effects and impacts on right holders, developers and users, including effects on SMEs and individuals.
- 133.** However, many right holders and creators who supported the current law also had concerns about their ability to enforce their rights, including where their works have been used in other countries. They sought the introduction of measures to help them do this more effectively, such as transparency measures.
- 134.** Many AI developers, and others from the technology and research sectors, did not think current copyright law strikes the right balance to support AI development and adoption in the UK. They felt that the restrictive nature of UK copyright law compared to that of other countries, such as the USA, meant that UK developers and users faced greater risks and costs, holding back development and use in the UK. Respondents from all sectors noted that challenges under the status quo can affect SMEs and individuals in particular, owing to their relative lack of time, resources, and legal knowledge, but large firms and other organisations also noted similar challenges.

135. Generative AI technology continues to develop rapidly, and the copyright licensing market is continuing to grow and take shape. There also remains uncertainty about how the regulatory environment will develop in other countries, notably the USA and EU, where significant legal cases remain to be decided and new regulation has only just begun to take effect.
136. The next section considers potential options to change copyright law, in light of views from the consultation and developments that have taken place since.

Section C: Copyright and AI systems – Consultation options

Overview

1. This section discusses the likely effect of potential changes to copyright law that were considered in the consultation.
2. The government wants to ensure that AI can fulfil its potential to grow the economy, increase productivity, and improve living standards, without undermining human creativity or our world-leading creative industries. In this section, we assess the likely effects of potential changes in light of this ambition and stakeholder feedback.
3. The consultation set out three objectives for copyright and AI development:
 - **control:** right holders should have control over, and be able to license and seek remuneration for, the use of their content by AI models.
 - **access:** AI developers should be able to access and use large volumes of online content to train their models easily, lawfully and without infringing copyright.
 - **transparency:** the copyright framework should be clear and make sense to its users, with greater transparency about works used to train AI models, and their outputs.
4. The government set out three potential options for changes to copyright law, as well as the status quo option. In this section, we consider first the preferred option at consultation stage – an opt-out exception – followed by the two other options – stronger copyright, and a broad exception without opt-out. Our consideration of these options is informed by responses to the consultation, as well as other sources. Potential economic impacts of each option on right holders, AI developers, and AI users are considered further in the accompanying impact assessment. We also consider the potential for more targeted interventions, based on ideas put forward in consultation responses.

A broad data mining exception, which allows right holders to reserve their rights by opting out (consultation option 3)

5. At consultation stage, the government’s preferred option was to introduce a new exception to copyright for “data mining”. The exception would apply only if a right holder had not expressly reserved their rights – or “opted-out” – through machine-readable means. This exception is referred to below as the “opt-out exception”. It would be similar to that provided in EU law by Article 4 of the Digital Single Market

Directive,²² and alignment with EU law was one of its potential advantages. It would have the following features:

Data mining

6. The opt-out exception would apply to “data mining”. Data mining is the use of automated techniques to analyse large amounts of data, in order to identify patterns, trends and other useful information. The UK’s existing data mining exception, which applies to non-commercial research only, uses the term “computational analysis”.
7. The general rationale for data mining exceptions is that they enable people to conduct analysis on, and gather information from, content that they have lawful access to, using tools that speed up processes that they otherwise would have to do manually.
8. Data mining exceptions allow users of automated tools to extract non-protected information (facts, trends, etc.) from protected works, and to make any necessary copies in the process, without breaching copyright in the work.
9. Under this option, the exception would permit data mining for any purpose, including the training of AI models.
10. Data mining exceptions typically apply only to the right to make copies and do not allow works to be communicated to the public.

Lawful access

11. The opt-out exception would only allow data mining to take place on works to which a user has lawful access. Examples would include works made freely available on the internet and those made available under contractual terms, such as via a subscription.
12. The reason for restricting the exception to lawfully accessed works is that, in principle, it allows the right holder (or their licensee) to factor data mining into any access conditions applied to the work, and what is charged for that access. The UK’s existing exception for non-commercial research data mining also has a lawful access restriction, as do the EU’s data mining exceptions.
13. Some respondents to the consultation noted that the term lawful access is “ambiguous”. Some from the research sector were concerned about conservative interpretations of it, which may unduly restrict access and make it difficult for them to carry out their research. However, many right holders felt that the term was clear and should be interpreted restrictively. In particular, they felt it was clear that it would not cover pirated material, and argued that it should not permit use of material which is licensed for one purpose (e.g. listening to music via a streaming platform) for the additional purpose of TDM.

²² 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

Rights reservation

14. The existence of an opt-out means that the exception can be “switched off” by right holders who expressly reserve their rights. For works available online, rights would need to be reserved by machine-readable means.
15. Under the EU’s exception with opt-out, a reservation must be made expressly by the right holder using “appropriate means”. This means a machine-readable means for online content. There are a variety of ways by which this may be achieved, and some uncertainty about which means can qualify. The government, in its proposal, said it would take a similar approach to the EU but would seek to provide greater certainty over valid opt-outs. Technical means for rights reservation are considered further in Section F of this report.

Exception with opt-out: views from consultation

16. This option was supported by 3% of consultation respondents. This included many from the AI, technology and research sectors, amongst others. Many of them recognised that this option presents a compromise between different interests.
17. The IP Federation, representing views from innovative UK industry, urged the government to enact this option “without delay” to “ensure the UK remains a leading jurisdiction for AI and technology investment”.
18. Tech UK, representing firms across the UK’s technology sector, said that some of its members supported option 3 as the “minimum bar necessary” for maintaining UK competitiveness in comparison to the EU and globally. But they argued that several areas need to be addressed to ensure that an “opt-out” approach worked in practice.
19. The Law Society noted that greater clarity of the law was needed and that “a clearly defined and scoped means for data mining where rights holders continue to reserve their rights, in line with EU requirements” could meet these objectives. RELX, a journal publisher and provider of analytical tools, said that “with the right safeguards” this option could “bridge the gap between the views of right holders and AI providers”. Adobe, a specialist in content creation software, noted that, while option 2 (a broad exception) would be the most attractive to AI developers, “it is also important to recognize and address concerns of creators, especially given the thriving creative sector in the UK”.
20. Some large AI developers also supported this option, or modified versions of it, on the basis of it being a compromise which sought alignment with the EU. Those who supported this option did so subject to balanced solutions on transparency and technical tools, and some proposed modifications.
21. For example, Google – in support of a modified option 3 – said they “believe rights holders should have choice and control” and noted existing technical solutions such as robots.txt that could achieve this. They supported this approach alongside a specific exception for research.
22. Many supporters of this option emphasised the need for effective rights reservation mechanisms before the exception is introduced. Several respondents noted that

such technologies are already in development or in active use. A detailed summary of feedback on technical solutions is provided in Section F.

23. However, by far the greatest number of consultation respondents opposed this option. Strong concerns were raised by the creative industries, including individual creators and performers.
24. Individual creators and performers, and groups representing them, emphasised the impacts on individuals and SMEs. A template response (submitted by a large number of individuals) said this option “would disproportionately harm small creators, who are the least likely to reserve their rights and the most vulnerable to being out-competed by AI models trained on their work”. The Creators’ Rights Alliance described the implementation of opt-outs by individual creators as “significantly burdensome”, and for large volumes of works “impossible”. Several right holder respondents warned that, without an easy opt-out mechanism, option 3 could be similar to option 2 – the broad exception – which they opposed.
25. The Motion Picture Association said, “A rights reservation mechanism would amount to a regulatory burden falling on a wide range of rightsholders, putting the onus on them to protect their rights while not all will have the resources to do so”.
26. As well as its practicality, many were opposed in principle to a copyright exception based on an opt-out. UK Music, representing parties from across the UK music industry, said it “undermines the permission-based principle of copyright: consent must be obtained, never assumed”. They said the proposed approach would lead to losses of millions in potential licensing revenue.
27. A number of AI developers also opposed this option, but for different reasons. In the view of several foundation model developers, ways could be found to make it easy to opt out of the exception, but the result would be that it ends up having little impact in practice. These organisations typically supported a broad exception – option 2 in the consultation.
28. Consultation questions 6 to 8 sought specific views on the legal approach to an opt-out. They asked what action developers should take when a reservation has been applied to a work, the legal consequences if it is ignored, and whether rights should be reserved in machine-readable formats.
29. The response by the UCL Institute of Brand and Innovation Law (authored by Dr Alina Trapova) highlighted the issues to be addressed, with reference to debates over EU law. These include the timing of a reservation (whether an opt-out may take effect when data mining has already taken place); the location of a reservation (whether it should be applied at the point where data is obtained, such as on a website, or can be notified on a universal basis, for example through a registry, or both); and whether reservations can be individual or general (applied per work or collectively, for example by a Collective Management Organisation).
30. Many right holders who commented on this aspect (without necessarily supporting this option) were in favour of being able to reserve rights through a wide range of methods and means. For example, the Alliance for IP said that any reservation must apply “irrespective of when, where and how [a work] might be copied”.
31. Right holders also supported the effects of an opt out going beyond the specific copy being used and the specific person the opt out is notified to. So, if a work is

opted out in one location, or registered as opted out in a database, the consequence would be that it could not subsequently be used for data mining by any person.

32. Some respondents raised challenges relating to content with overlapping rights and co-authored and derivative works, noting that if one collaborator were to opt-out but others did not the status of the opt-out may be unclear.
33. Many AI developers were concerned about a broad approach to an opt out. They favoured specific forms of rights reservation, in particular location-based signals which can be read by web crawlers. They also raised concerns about opt-outs that could be invoked across all users and at any point, favouring opt-outs that apply at the point where data is acquired.
34. For example, Open AI were concerned that “each new training run would require ongoing legal reviews to verify whether newly reserved rights impact their training data, creating a cycle of verification, removal, and retraining”. Knowledge Rights 21, representing views from innovation and research sectors, said this approach would require people to “relicense access to content they already have lawful access to”. They said this “will introduce huge overheads and reduce access to data”, making the UK less competitive than other economies.
35. In summary, on the legal effects of a rights reservation approach, there was little common ground. The creative industries tended to support broad opt-outs via a range of means, though noted complexities associated with this. For many of them their strong preference was no exception at all. The AI sector supported specific opt-out methods at clearly defined points, though many preferred an exception without an opt-out.
36. There was also a range of opinion on technical measures and standards that would implement any opt-out. These are discussed in detail in Section F of this report. They are becoming increasingly sophisticated and will play an important role in access to, and use of, copyright works with AI systems. But, although consultation respondents tended to support their use, there was little consensus as to their efficacy, and ease of use, in the context of an opt-out approach.

Opt-out exception: potential effects

37. Were it to be introduced, an opt-out exception is likely to have the following effects on the use of content with AI systems, compared to the status quo:

Data acquisition

38. An opt-out exception is likely to make it easier, to some extent, for data aggregators to assemble datasets in the UK for provision to AI developers. In *Kneschke vs. LAION*, the Hamburg courts²³ have found that the EU’s opt-out exception – which the option 3 exception resembles – would cover copies made while analysing (lawfully accessible and non-reserved) works and assembling them into a dataset, as preparatory acts to AI training. As such, we expect an opt-out exception would apply to certain activities performed by dataset providers, where they meet the relevant criteria.

39. Permission would still be required to communicate works to the public, as the data mining exception would only apply to the making of copies. However, dataset providers do not always directly communicate works but may instead provide links to works already available to the public.²⁴

Pre-processing and training

40. The exception would also permit an AI developer to download copies of lawfully accessed works for the purpose of training a model, as long as the work has not been opted out by a right holder. This could benefit AI firms that want to train their models in the UK. The size of the impact on AI developers will depend on the extent to which works are opted out of the exception, and the nature of those works.
41. As noted above, the “pre-training” stage typically requires a very large number of works – often in the billions. If a large number of works (in general, or of specific types) are opted out, then it is unlikely that AI developers will be able to access a sufficient volume of works under the opt-out exception. This was the scenario raised in consultation responses by several AI firms which opposed this option. While larger firms may be able to supplement works available under the exception with other sources, this may be less feasible for SMEs, individuals and startups.
42. In that scenario, the impacts may not be that different to the status quo. A developer seeking to train a model in the UK will typically require licences from right holders. They will be able to access a certain number of works under the exception, but to obtain a sufficiently large and generalised dataset they will need to secure licences. This will be particularly true in relation to high value works which are more likely to be opted out.
43. An alternative scenario is that argued by many creative industry respondents to the consultation. Under this scenario, many works are not opted out, in part because of the difficulty of doing this, and the range of platforms on which copies may appear (see the technical standards section for more discussion of opt-out feasibility). If it is legally or technically challenging to opt out, this is more likely to affect SMEs and individual right holders who have fewer resources than larger right holders, meaning more of their works are available.
44. In the latter scenario, the impacts in the UK would be more akin to those under option 2 – the broad exception. AI firms would be able to access large datasets for LLM training under the exception. They may also supplement these with licensed datasets, either at pre-training or fine-tuning. Licensing is likely to remain the main way that developers obtain content that is not publicly available online. But they would also be able to use a lot of material without expressly licensing it.
45. In this scenario, right holders argue that it will be harder to license their content for AI training in the UK, as it would be “competing with free”. They would also face greater competition from the outputs generated by AI models, without compensation for use of their works by those models.
46. At this stage, as there remains significant uncertainty as to how many works would be available under an opt-out approach, it is difficult to forecast potential impacts.

24 For example, the LAION datasets.

Choice of training jurisdiction

47. Potential impacts are further complicated by the fact that most foundation models are not trained in the UK. Many are trained in the USA or China, for example. The extent to which works are easily available for training in the UK may affect the likelihood of developers choosing to locate their training here, alongside other factors.
48. If a large and diverse number of works are available for use under an exception the UK may become a more attractive place for training to take place. If that is the case, we would expect more AI training to take place in the UK.
49. However, even if this is the case, AI developers will continue to be able to rely on other countries' laws, including the laws of the USA and EU, if they conduct their training in those countries. A developer is able to choose which jurisdiction they train their model in. They can take this decision based on a comparative analysis of the regulatory environment in different countries, including copyright, as well as other factors such as support for startups and SMEs, energy costs, and compute infrastructure. The attractiveness of any new exception in the UK, therefore, will be affected by developments in other countries, including the outcome of fair use litigation in the USA, and wider structural factors.
50. UK-based firms as well as foreign ones are able to conduct their model training outside the UK. The judgment in *Getty Images vs Stability AI* (which is currently subject to appeal) describes such a situation. As such, it may be small and micro firms, which are less able to manage the cost and legal risk of training overseas, who wish to rely on a UK exception rather than larger firms.
51. Given the uncertainty over the extent to which opt-outs would be exercised, and their effects, and the many other factors that AI firms consider when choosing where to locate their training activity, the introduction of an opt-out exception may not be decisive in many AI firms' choice of training location. Firms already based abroad may see little reason to transfer their activity to the UK, and some have chosen to train overseas using licensed UK datasets. The Startup Coalition, in its consultation response, said that AI models "will still be trained on UK data whether UK rightsholders have opted-out or not, the only difference will be the UK will not see the economic upside".

Model release

52. The opt-out exception is unlikely to have significant impact at the point of model release in the UK, compared to the status quo. The exception would apply to the reproduction right, and would only apply to data mining, or computational analysis. It would not affect other rights, such as the right to communicate works to the public. It would also not affect secondary infringement for acts such as importing, possessing, and dealing in infringing copies.
53. However, for such rights to be engaged, a copyright work needs to be present. As noted under the description of the status quo, a trained AI model or system might not contain any copies of works, or communicate them in its outputs.
54. To the extent that works are present in the model or its outputs, the existing law on communication and distribution to the public, and secondary infringement for dealing with infringing copies, would apply.

Task-specific training and development

Fine-tuning

- 55.** This option is likely to have similar effects for AI developers who take a trained model and fine-tune it as it would for those who train a foundation model. A model will be fine-tuned for a specific application. As such:
- The size of the dataset used for fine-tuning is likely to be smaller than a dataset used to train the original model, though may still be quite large.
 - The required dataset may be specialist, so not easily obtainable through means such as web crawling.
 - The person doing fine tuning for a UK-based application may be more likely to be based in, and make copies in, the UK.
- 56.** As discussed above, an opt-out exception is likely to make it easier for AI developers to train models in the UK, by reducing their risk of copyright infringement. However, this would only apply to the extent that works are not opted out of the exception. The extent to which AI developers are able to benefit from the exception at the fine-tuning stage will therefore depend on various factors, including the extent to which opt-outs are implemented for the type of material they are trained on, the general accessibility of that material and the size of datasets required.

Retrieval-augmented generation (RAG)

- 57.** The exception would also apply to material used in RAG, to the extent that this involves copying and analysing works. An example could be the use of up-to-date content available on the internet to provide answers by an AI assistant, or the provision of information from web pages by an AI search engine.
- 58.** Many providers of high value and high information content on the internet, such as news media sites, social media platforms, and web forums, already seek to block access to certain web crawlers, through means such as the robots.txt standard (described in the section on technical standards). Some may choose to block certain crawlers and permit others, depending on the purpose of those crawlers. Such measures would be considered valid opt-outs under this exception. We would seek to encourage greater standardisation, as well as adherence to standards by web crawlers and AI developers.
- 59.** As such, a certain amount of – potentially high-value – online content would be likely to be opted-out of this exception for the purpose of RAG. However, as noted above, many right holders were not convinced that opting out via technical means would be effective, or would be respected by AI developers. Such views were strongly held by many in the news media, whose content is often seen as high value in terms of RAG.
- 60.** There is a degree of uncertainty, therefore, as to how much impact this exception would have on RAG. As discussed under AI training, a range of scenarios are possible. In one scenario, few works would be opted out (or opted out effectively) thus allowing a wide range of content to be available. In such a situation, it would be difficult for right holders to secure licences for RAG purposes. In another scenario,

many works would be opted out, thus requiring many high-value sources of information to be licensed for this purpose.

Deployment

- 61. The opt-out exception could also have an effect on the deployment of AI applications and the use of AI tools in the UK.
- 62. The reduced legal risk to developers of AI systems could mean that a wider range of tools are provided in the UK than in the status quo scenario, and the cost of providing and deploying such tools may be decreased.
- 63. There could also be reduced risk to firms deploying AI in their organisations – for example, to analyse business and other data contained in third party material.
- 64. The size of these potential effects will depend on the extent to which works are opted out of the exception. The extent to which this would occur is unclear, as are the effects of this exception on different categories of user.

Exception with opt-out (consultation option 3): assessment

- 65. We want to ensure that our approach to copyright allows us to realise the extraordinary potential of AI to grow the economy, create new, more rewarding jobs, and improve living standards, while ensuring our creative industries continue to flourish in the age of AI.
- 66. There was strong opposition to the proposal for an exception with opt-out in the consultation. This came from a range of sectors, but most significantly from the creative industries. They were concerned that this approach would lead to their works being used to train AI models against their wishes and would reduce their ability to seek remuneration for it.
- 67. Many technology developers predicted that it would lead to high take-up of opt outs, undermining the purpose of the exception to support access to works. They were concerned this would make it more difficult to develop their models in the UK compared to other jurisdictions.
- 68. We said we would not take forward this approach unless we can be confident that it will meet our objectives. But there are still significant gaps in the evidence base – including in terms of how copyright impacts the wider development and deployment of AI – and uncertainty about the impact of reform.
- 69. There are also ongoing international developments that we are monitoring. The proposal for an opt-out exception is similar to the approach taken by the EU. The EU exception continues to be considered by its courts, which will help to give greater clarity over its scope and effects.
- 70. In view of the concerns raised by stakeholders, and the continued uncertainty about the likely effects of an exception with opt-out, a broad copyright exception with opt-out is no longer the government's preferred way forward.

No new exception, copyright is strengthened so

licensing is required for AI development (consultation option 1)

71. Option 1 in the consultation was to strengthen copyright so that licensing is required for copies made during AI development in all cases. It was supported by a majority of respondents to the consultation: 81% of respondents stated option 1 as their preference. As many respondents noted, UK copyright law already provides extensive rights, with limited exceptions. As such, under this option copyright law would remain largely the same as the status quo.
72. Many respondents in favour of this option were concerned that – although broad rights were available under UK law – it was not as easy to exercise these rights as they wished. They were keen to promote licensing through easier enforcement of existing law, in particular through implementation of robust transparency measures. Input transparency measures are discussed in Section D of this report. Additional measures that are relevant to this option also include technical standards, licensing and enforcement, which are covered in later sections.
73. Under this option copyright would broadly be the same as the status quo. However, the consultation did seek views on certain potential changes that might make copyright more restrictive. These were:
 - Clarifying existing exceptions – in particular the temporary copies exception and the non-commercial research data mining exception.
 - The treatment of models transferred to the UK from other countries.

Existing exceptions

74. Existing exceptions may be relevant to the training and development of AI models in the UK. In particular, the consultation highlighted the temporary copies exception (section 28A CDPA) and the data mining exception for non-commercial research (section 29A CDPA). The consultation asked whether the former should be clarified (questions 26 and 27) and whether the latter remains fit for purpose (question 28).
75. The temporary copies exception permits temporary copies to be made during technological processes – for example, copies held in browser caches or displayed on computer screens. The exception – which derives from EU law – is intended to ensure that such processes can function efficiently.
76. Many respondents to the consultation noted that several conditions must be met before this exception can apply. Any copies made under it must be transient or incidental, and an integral and essential part of a technological process. They must be made solely to enable transmission in a network, or a lawful use of the work, and they must have no independent economic significance.
77. The data mining exception for non-commercial research was introduced into UK law in 2014. It aimed to make it easier for researchers to use modern data mining techniques, allowing them to analyse large quantities of research text and uncover patterns and trends, supporting scientific discovery.
78. Many respondents to the consultation also noted the limited scope of this exception. The exception applies only to non-commercial research and requires a sufficient

acknowledgement to be made of the source, where practicable. For example, the Centre for Regulation of the Creative Economy (CREATe) based at the University of Glasgow highlights its limitations for the purposes of AI research particularly: “While the existing exception can be interpreted as permitting certain AI training activities, it does not cover the whole lifecycle of AI models and systems. Together with other legal regimes, it creates uncertainties at all stages of AI research and development.”

79. Under option 1, which aims to strengthen copyright, these exceptions could potentially be clarified or amended to ensure a narrow application.

Models trained in other countries

80. The consultation asked to what extent the copyright status of AI models trained outside the UK requires clarification, to ensure fairness for AI developers and right holders. It said the government would consider whether action could be taken to help establish a level playing field between providers of models trained in the UK and those trained outside the UK but made available for use in the UK market.
81. The consultation did not put forward any specific proposals for reform to copyright in this area. However, some respondents to the consultation said that secondary infringement provisions in UK copyright law do, or should, apply to models trained in other countries. For example, the Publishers’ Association said that where “an AI developer has trained its model outside of the UK, but on content protected by UK copyright, it follows that making that model available in the UK amounts to copyright infringement”.
82. This issue has been the subject of litigation in *Getty Images v Stability AI*. As noted in the previous section, the court said in its judgment that an AI model trained outside the UK could be considered an “article” under copyright law, so its importation could be a secondary infringement under section 22 CDPA. However, to infringe under this provision the importer would have to know or have reason to believe that the model is an infringing copy. In the judgment, the particular model was held not to contain copies of the relevant works. This judgment, and its interpretation of this provision, is currently subject to appeal.

Stronger copyright: views from consultation

83. A wide range of individual creators, creative firms small and large, and groups representing creative industry sectors, supported stronger implementation of existing copyright law. This included the Alliance for IP, the British Copyright Council, the Creators’ Rights Alliance, UK Music, the Publishers’ Association, the Motion Picture Association, and the News Media Association, among others.
84. As noted above, the main focus of these responses was on maintaining the copyright status quo, as the law already provides extensive rights. Their support for option 1, as opposed to the status quo, was based on the introduction of transparency measures. Input transparency measures are discussed in Section D of this report.
85. As the Financial Times stated: “There is no conflict between promoting AI innovation and upholding copyright. It is already evident AI developers and the creative industries need each other in order to succeed. The existing copyright regime ... is

central to driving a virtuous circle of investment of human made creative material and ensuring that material is used in line with the wishes of the creator”.

86. This view was shared by many right holders in the creative industries. For example, the Publishers’ Association cited a report asserting that AI developers currently spend only 0.1% of their budgets on licensing, noting the potential for that to grow significantly if copyright is more easily enforced.
87. Many responses supporting option 1 also considered that an exception with rights reservation would be impractical, or unacceptably burdensome for the creative industries. This was a view particularly strongly noted among organisations representing individuals and smaller businesses.
88. Regarding existing exceptions, few respondents sought additional restrictions to either the temporary copies exception or the existing non-commercial research exception. This was often on the basis that the exceptions were already narrowly drafted.
89. For example, the British Copyright Council said the requirements in the temporary copies exception “make it clear that the exception is not a replacement for the licensing of restricted acts involved in the commercial crawling and database compilation and storage of works linked to generative AI model developments”. The Alliance for IP agreed and sought clarification of this through a formal copyright notice.
90. Some respondents did seek such clarification to be made in law. Some made these comments in relation to the option 3 opt-out exception saying that, if it went ahead, it should be clear that data mining could be done only under that exception and not others.
91. Some respondents expressed concerns about the potential misuse of the existing non-commercial research data mining exception. For example, RELX suggested that guidance on the exception should “make it clearer that prior to the subsequent commercialisation of a model that has been trained using works under the non-commercial exception the licensing of works from the original copyright owners will be required”.
92. Technology firms and researchers generally warned against reducing the scope of these exceptions, and some called for their expansion. Views on broadening exceptions for research are considered under the “targeted exception” option, below.
93. Regarding models trained outside the UK, several right holders stated that UK copyright already addresses this situation and that it is an infringement to import into the UK a model trained outside the UK, which would have infringed if trained under UK law. This issue should now be viewed through the lens of the judgment in *Getty Images v Stability AI*, which was issued after the consultation period, and which is subject to appeal.
94. Some sought non-legislative clarification of the law on imported copies. For example, the Alliance for IP suggested the IPO issue a copyright notice setting out the position.

95. Other right holders called for amendments to copyright law, to provide that models trained outside the UK, under the copyright law of their training jurisdictions, should also comply with UK law at the point of deploying in the UK market. However, many proposed a regulatory approach to the importation of such models, similar to that taken by the European Union in its AI Act, rather than through changes to copyright law itself. These respondents argued that this would lead to increased licensing, as developers would not be able to rely on exceptions in other countries when developing models for the UK market.
96. On the other hand, respondents from the AI and wider technology sector warned against such amendments to copyright law, noting that they could have a significant impact on the availability of models in the UK. For example, the IP Federation urged the government not to introduce such measures, noting that the best way to achieve a level playing field was by creating a favourable environment for AI development in the UK.

Stronger copyright: potential effects

97. If copyright were strengthened, as described above, this could lead to increased licensing, as developers would find it more difficult to rely on exceptions (either in the UK or in other countries) when developing models for the UK market. Consequently, this could lead to increased remuneration for copyright owners in the UK. As discussed in the licensing section of this report, it is unclear to what extent the benefits of licensing deals will flow to SMEs and individuals rather than larger right holders.
98. If UK copyright law was applied to the training of models trained abroad when put on the UK market, this would affect the providers of those models to the UK, and downstream developers and users of them. Each may require licences depending on their activity (for example, importing a model or dealing with it). The effects on right holders, developers, and users, are likely to be the largest in relation to foundation models and other models currently trained in other countries. There are likely to be fewer effects on training and development that takes place entirely within the UK and which already requires licensing.
99. While this approach could lead to some models trained outside the UK being provided on a licensed basis, there is also a risk that certain models and systems, including foundation models, will not be provided to the UK market, where the costs and risks of doing so outweigh the benefits to the provider. Where new models are provided to the UK market, they could be delayed or have restricted features.
100. If this were to happen, it could affect a wide range of downstream developers of AI systems that are built on foundation models, as well as users of AI models and systems developed abroad. This could reduce the potential productivity benefits of AI across the economy. Consultation responses from UK developers indicated that developers who build on foundation models in the UK are often startups and other SMEs. Tools and services built using these models are employed by businesses of all sizes, as well as other organisations and individuals, across the UK. Therefore, the effects, while uncertain, could have wide-ranging consequences across different economic sectors.

Stronger copyright: assessment

- 101.** The strong feedback to the consultation, particularly from the creative industries, was that copyright works should be licensed for AI training in all cases.
- 102.** Many did not think this required amendment to existing UK exceptions, as copyright already applies to various acts in the training process, and existing exceptions are limited in scope. Although some right holders called for amendment to existing exceptions, others felt that their limits were clear. Restricting them further could also have unintended consequences for the technology firms and researchers who rely on them for many reasons unrelated to AI. Based on consultation views, we do not think these existing exceptions are a significant obstacle to licensing in the UK, and licensing is more likely to be encouraged through other measures, such as greater transparency.
- 103.** AI is increasingly integrated into our lives, and its benefits are being realised by many different parts of the economy. The AI tools that people use daily are often developed – at least in part – in other countries. Intervening to restrict the supply of these models to the UK could have significant impacts for users of AI in the UK, as well as UK developers and innovators. While some AI providers may seek licences for training that has already taken place in other countries, others may choose to withdraw from the UK market altogether. This could have significant impact on people’s ability to use this technology in the ways they have become increasingly used to. Other interventions, such as statutory licensing (see Section G) for the use of such models, are also untested, with unclear impacts and implementation challenges.
- 104.** Since the consultation, the *Getty Images v Stability AI* judgment has provided greater clarity over the law in this area. However, this judgment is currently subject to appeal. We believe it is right that the application of the law in this area should be considered properly by the courts, to give certainty to right holders, developers and users of AI models.
- 105.** Developments in other countries will also have a significant effect on outcomes in the UK. We are encouraged by the increasing amount of licensing taking place in other jurisdictions, including of content created and owned by UK creators and businesses. We expect to see greater clarity over the scope and effect of the law in the USA and EU in the coming year. There may be some levelling of the playing field internationally as laws are clarified or changed and licensing agreements are reached. These developments could lead to greater licensing of UK content across national boundaries, without domestic intervention.
- 106.** As such, we propose that we do not amend copyright law in respect of systems developed outside the UK at this time. This means that the law as interpreted in the *Getty Images* case will continue to apply. We propose to keep this issue under review, taking account of legal outcomes in the UK and elsewhere, and wider market developments.
- 107.** We also propose to engage with other countries, including bilateral engagement and through forums such as the World Intellectual Property Organization (WIPO). We propose to further strengthen our international cooperation on AI and copyright, with the aim of supporting closer alignment between countries.

A broad data mining exception, with no opt-out (consultation option 2)

- 108.** Option 2 in the consultation was to introduce a broad data mining exception. This would allow data mining on copyright works, including for AI training, for any purpose and would be subject to few or no restrictions. It was supported by a very low number of respondents.
- 109.** The consultation did not define a broad data mining exception in detail, but cited exceptions in other countries, including the USA, Japan and Singapore. It should be noted that a number of respondents criticised the framing of these exceptions as broad, noting that they did, in fact, contain a number of restrictions.
- 110.** For example, Japan’s data mining exception does not apply to uses that would unreasonably prejudice the interests of the copyright owner. It is also limited to purposes that do not allow someone to enjoy the thoughts or sentiments expressed in the work. The Libraries and Archives Copyright Alliance referenced an academic article which describes the Japanese exception: “Similar to US fair use, the two Japanese exceptions create an open list of permitted activities and apply a uniform flexible test to their application. Nevertheless, unlike fair use they create a codified outer limit to their application and purpose.”²⁵
- 111.** The USA fair use principle is subject to four factors as outlined previously. Right holders argued that these disallow many uses currently made by AI developers (though AI developers argued the opposite). The ultimate scope of these exceptions will be subject to litigation in those countries, and it is difficult to assess their likely effects. However, it may be that they end up being found to have narrower scope than the broad exception proposed in the consultation. Some right holders, taking the view that fair use will be interpreted narrowly, said that following the USA would be preferable to following the EU.
- 112.** Therefore, for the purpose of our analysis in this report, we model the broad data mining exception as having no principle-based constraints such as “fair use”. This is akin to the exception available in Singapore.
- 113.** An exception of this type could have the following features:
- a) Data mining**
- 114.** Like the opt-out exception, the broad exception could permit “data mining” for any purpose, including the training of AI models. Data mining exceptions allow users of automated tools to extract non-protected information (facts, trends, etc.) from protected works, and to make any necessary copies in the process, without breaching copyright in the work.
- 115.** It would be an exception to the right to make copies but, like other data mining exceptions, it would not apply to other rights attached to copyright.

²⁵ Mendis, D., White, B., & Hong, D. (2024). Copyright and Open Norms in Seven Jurisdictions: Benefits, Challenges & Policy Recommendations. Zenodo. <https://doi.org/10.5281/zenodo.10655757>, Available at SSRN: <https://ssrn.com/abstract=4728782>

b) Lawful access

- 116.** The exception would only apply to works to which the person doing the data mining activity has lawful access. Examples would include works made freely available on the internet and those made available under contractual terms, such as via a subscription.

c) No opt-out

- 117.** Unlike the option 3 (opt-out) proposal, the broad exception would not allow people to reserve their rights.

Broad exception: views from consultation

- 118.** The broad exception had the lowest level of support of the consultation options and was overwhelmingly opposed by creative industry respondents. However, broad exceptions were supported by some of the largest providers of AI foundation models and services. Several encouraged the government to consider a more flexible approach than the opt-out proposal, akin to those in the USA and Japan.
- 119.** Those who supported this option emphasised its impact on potential investment in the UK. OpenAI said they were able to invest in the USA because “U.S. copyright law has exceptions including fair use that protect AI development. If the UK adopts a straightforward copyright regime, AI businesses will similarly have the legal certainty necessary before investing billions of pounds in long-term infrastructure and technology development.”
- 120.** Supporters also focused on wider benefits of this option. Anthropic said, “Option 2 is consistent with copyright’s purpose and function. It will provide the legal certainty to not only develop more advanced AI models in the UK, but also improve AI safety, reduce bias, and foster robust competition.”
- 121.** Smaller UK-based AI developers also supported this option. The Startup Coalition, representing early-stage British technology companies, said it was the only option “that would align the UK’s position more closely with other countries also going for gold in the AI race”. They argued that other options would entrench asymmetries between startups, which are often small and micro businesses, and the “tech giants”.
- 122.** However, the broad exception was strongly opposed by creators and the creative industries, for similar reasons to their criticism of option 3. Many expressed concern that such a broad exception would mean AI developers could use their works with AI, without recompense, to produce outputs that compete with their own works.
- 123.** Sony Music Entertainment noted that without a functional opt-out, option 3 is the same as option 2. Option 2 is opposed strongly. “Option 3, like option 2, would therefore deprive right holders of copyright protection for data mining and AI training, which would cause severe economic damage to the creative industries, disincentivise artists from creating content and contributing to UK culture, and significantly reduce the creative industries’ contribution to the UK economy.”
- 124.** Some argued that a broad exception which has these effects would be difficult to justify under the “three-step test” provided in international law. The British Copyright Council opposed option 2 alongside option 3, specifically highlighting international

obligations: “The exception outlined in Option 3 (and also Option 2) contravenes the internationally binding Berne Convention Three-Step-Test... Options 3 and 2 infringes [sic] every one of the accumulated clear applicable steps as defined for instance by the WTO Dispute panel (case WT/DS160/R).”

Broad exception: potential effects

Data acquisition

- 125.** As noted under option 3, a data mining exception could make it easier to assemble training datasets, by allowing copies to be made during their assembly. The compiler of a dataset would be able to perform data mining on works lawfully made available on the internet, regardless of whether a right holder had sought to reserve their rights. This could make it easier to compile unlicensed datasets in the UK. As this version of an exception would not include an express opt-out, we would expect more works would be available for use in such datasets, which could make it more difficult for right holders to enter into licensing agreements for these uses, or to license their own datasets.
- 126.** However, as with other options, the exception would only apply to the right to make copies. So, while a dataset provider may rely on it for analysis – for example to match metadata to images – it would not be able to rely on it to distribute or communicate works to the public.
- 127.** The availability of this exception is likely to have a corresponding effect on the provision of licensed datasets. Should it become easier to provide unlicensed datasets we would expect fewer developers to seek to use licensed datasets (to the extent that such datasets are substitutable). This could affect the provision of licensed datasets in the market.
- 128.** It could also influence more right holders to put their content behind paywalls or use other technical means to restrict access, leading to a reduction in freely-accessible information online.

Pre-processing and training

- 129.** The broad exception would also permit an AI developer to download copies of lawfully accessed works for data mining while training a model. This would benefit AI firms that train their models in the UK, in particular those which seek to train using online materials. Given the current practices of foundation model developers, and their responses to the consultation, we would expect them in particular to benefit from this exception. This is because of the sheer volume of data points foundation models need to be trained on at the “pre-training” stage, which is commonly done using online material.
- 130.** With no opt-out, the broad exception would allow access to a larger quantity of works than option 3. The overall effect would be easier access to material by AI developers, but reduced licensing opportunities for right holders.
- 131.** However, right holders are still likely to seek to employ means to prevent access to and use of their works, such as by blocking web crawlers, meaning not all online material will be available for AI training. We would expect the prevalence of such

measures to increase. There may also be reduced incentives to create certain types of content, leading to a reduction in its availability for AI training.

- 132.** Despite the restrictions they may put in place, we expect this option to have the greatest potential effect of any option on right holders by reducing their ability to license their works for training in the UK. This would include high value works, which may be used to train AI models that directly compete in the same market as the originals.

Choice of training jurisdiction

- 133.** A broad exception could be more permissive than the EU's exception and broader than the fair use approach taken in the USA (depending on the outcome of litigation in the USA). As such, there could be a stronger incentive for AI developers to locate their model training (though not necessarily other functions) in the UK.
- 134.** If unlicensed training were to take place in the UK as a result of this exception, there could be a corresponding effect on licensing in other countries. Right holders in those countries may find it more difficult to agree licences for their works, as unlicensed training could take place in the UK. However, this is only likely to be the case if opportunities for unlicensed training were to be restricted in jurisdictions where it currently takes place, such as the USA.

Model release

- 135.** The broad exception outlined above would apply only to the right to make copies, and not any other right. As such, AI developers whose models or systems include copies of works would still need to comply with other aspects of copyright law, such as the right to communicate copies to the public (section 20 CDPA) and may also be liable for acts of secondary infringement, if their models are capable of outputting copies of works.

Task-specific training and development

- 136.** The broad exception would also have an effect on the further fine-tuning of models after their release on the UK market, as well as the integration of RAG functionality into services such as AI assistants.
- 137.** For fine tuning, the broad exception would increase the amount of training data available to developers without a licence in the UK. This in turn is likely to make it easier for them to develop new products and services, particularly if they are startups and other SMEs with limited access to resources.
- 138.** This will only be true to the extent that such developers use publicly-accessible content, such as online content. Where fine tuning relies on specialist datasets which are not publicly accessible, the exception would have limited impact. However, the overall effect is likely to be that more content is available to developers without a licence, with correspondingly fewer licensing opportunities for right holders in the UK.
- 139.** More data would also be available for RAG – for example in AI-enhanced search or by AI assistants. Opt-outs from RAG would not be formally recognised in this approach, meaning that developers building RAG-based services would not need to implement the same opt-out compliance architecture as in option 3. However, this

approach would still require works to have been accessed lawfully, and many sites will seek to block web crawlers, restricting the availability of works in practice. The extent to which this approach will affect RAG is therefore unclear.

Broad exception: assessment

- 140.** This was the preferred option of many in the AI sector, including large foundation model developers as well as UK startups. It appears to be the most favourable outcome for developers whose aim is to acquire access to large quantities of data at low cost, and the most likely to encourage foundation model training in the UK. However, support for this option was limited in terms of number of consultation responses overall.
- 141.** This option presents the greatest risks to revenue of right holders who have commercial licensing agreements in place or seek to enter into such agreements. It would also mean creators have less control over how their work is used, regardless of whether they are remunerated for it. However, the size of these effects would depend in part on whether model developers choose to train their models in the UK and how far content could be used in other jurisdictions with more permissive regimes. This approach could also lead to changes in behaviour, with more content becoming restricted by technological means, and reduced incentives to create and share new information. This could have consequences for the wider economy.
- 142.** A large number of developers who supported a broad copyright exception cited fair use in the USA and some suggested we should take a similar approach. But many right holders took the view that fair use permits a much narrower range of activity. The ongoing uncertainty about the scope of exceptions in other countries makes it difficult to assess the relative impact of this approach, were it adopted in the UK.
- 143.** Given the concerns raised by the creative industries, and the still uncertain international picture, we propose not to take forward this option at this time.

Alternative approaches

- 144.** Several consultation respondents proposed alternative options to those put forward. Some of these were in response to consultation questions on targeted approaches to copyright reform, including research exceptions, and whether rules should consider factors such as the purpose of an AI model, or size of AI firm (questions 28 and 29).
- 145.** These proposals often sought to strike a balance between the interests of creators, developers and users of AI, but in a different way from the options proposed in the consultation

Views from the consultation

- 146.** Several responses from the academic and research sectors emphasised the use of AI as a tool for furthering knowledge and innovation. The Centre for Regulation of the Creative Economy (CREATe) proposed a focus on research and development before market entry, with licensing and transparency being required at the point of market entry. Others focused on analytical uses of AI. The Bioindustry Association noted a range of benefits of AI for research, therapy and diagnostics in the life

sciences, and said that, once lawfully acquired, the rights to analyse information *“should be the same whether that information is read by a human or a machine”*.

147. Professor Emily Hudson (University of Oxford) and Dr James Parish (King’s College London) suggested focusing on reforms supporting “the genuine advancement of human knowledge” rather than “look-alike essays ... and images”. They suggested this could be done by clarifying or expanding the data mining and fair dealing exceptions which currently apply to non-commercial research. The Libraries and Archives Copyright Alliance also suggested that existing exceptions should be expanded. In particular: “the [section 29A non-commercial research data mining] exception should remain available to everyone, and focus on a broader research purpose that can capture how R&D will evolve... This is to reflect current research practice, including knowledge transfer, public-private partnerships, peer-review, accuracy checking and experimentation.”
148. Similar distinctions were drawn by other respondents. For example, one AI foundation model developer proposed an exception for commercial research, alongside the option 3 exception, separating out commercial use generally from research with commercial application. Another supported an exception which drew a distinction between generative and non-generative uses, with stricter conditions on generative uses.
149. Some organisations in the field of broadcasting and content distribution also supported some flexibility around non-competing and task-based uses. For example, one suggested an exception that applies to “task-focused” AI, rather than AI that generates new content. An advocate for creators’ rights said that if the government is not willing to maintain the current approach, it should consider laws that take into account the purpose of a model and its effect on the market. Some right holders suggested that (while not their preferred option) a USA “fair use” approach was preferable to an EU opt-out approach, as they expected it to be interpreted narrowly.
150. Some public bodies have also raised concerns about potential copyright infringement risks when using AI analytical tools to carry out tasks in the public interest.
151. As well as responses focusing on research or analytical uses, a smaller number of respondents referenced a levy system to provide payments to creators and right holders. For example, the Centre for Intellectual Property Policy and Management (CIPPM) at Bournemouth University, noted a broad exception with an AI system levy “ensures that AI system developers can progress with research and innovation, without being burdened by additional licensing, high transactive costs and practical complexities at the outset, whilst at the same time respecting human creativity”. Additionally, Dr Nicola Searle (Goldsmiths University of London) considered copyright levies more broadly and noted the potential for a levy to come from “advertising revenues (when AI generated content is supported by advertising), subscriptions or retail prices.” A levy system could be formulated in a number of different ways. As noted in Section B, on copyright in other countries, the Government of India is currently considering a proposal for statutory licensing that would create a broad TDM exception that requires AI developers to pay royalties for

the use of copyright works when AI systems are commercialised. Statutory licensing and levies are also discussed in Section G on licensing.

- 152.** As noted in our assessment of option 3 (exception with rights reservation) and option 2 (broad exception) we recognise creator concerns about a data mining exception which covers any commercial use and may compete with their own works. We also acknowledge that AI tools have a range of purposes and applications other than generating “lookalike” content. Some of these may support activity in areas of UK economic strength, identified by the government as having the greatest growth potential²⁶, such as science and research, or other areas of public benefit.
- 153.** Therefore, the final part of this section considers alternative approaches to the consultation options, based on suggestions put forward in response to the consultation.

Possible alternative approaches

A focused exception

- 154.** A focused exception targeted at certain types of use or application was the most popular alternative option suggested in the consultation. There were different views on what it could focus on, but many emphasised real world benefits to everyday citizens – supporting research, enhancing national security, and addressing other public policy goals – and also distinguished between analytical and generative uses. For example, a focused exception for science and research could make AI-driven scientific research easier, accelerating the discovery of new medicines, or supporting advances in fields such as climate modelling.
- 155.** Taking into account the different consultation responses, potential ways to approach a focused exception are outlined below. However, parameters would need to be more clearly defined and further evidence would need to be gathered before the impacts on right holders, AI developers, and users could be properly assessed.

Research

- 156.** While a number of possibilities were expressed by respondents, one approach to a focused exception could permit data mining, but only for the purpose of science and research. Restrictions could be included to minimise impacts on creators. It could be framed to permit copying to develop or use a model as an analytical tool. It could also be framed so it does not permit generation of content that competes with copyright works – meaning it could not be relied on, for example, to copy musical works for a model that generates new music that competes with the music of creators.

Public interest

- 157.** A focused exception could also apply to certain uses in the public interest, such as online safety and security. For example, AI tools can be used to identify harmful

²⁶ As identified in the UK’s Modern Industrial Strategy, https://assets.publishing.service.gov.uk/media/69256e16367485ea116a56de/industrial_strategy_policy_paper.pdf

content online. To do this they need to be able to ingest and analyse all relevant content. An exception focused on the use of data in the public interest could enable outcomes such as this.

Lawful access

- 158.** We would expect any exception to apply only to material which is lawfully accessed. This is the approach taken to existing data mining exceptions in UK and EU law, and consultation respondents supported similar restrictions. A focused exception could clarify this term, and help ensure right holders can manage access to their works, through a contract, subscription, or other means.
- 159.** This could be done by being clear that lawful access does not cover the use of pirated copies (for example, copies uploaded to the internet without permission). Right holders would also be able to control access to copies online using technical tools. Good practice on technical standards could help to define lawful access, to help give legal certainty to both developers and right holders.

A broad exception with remuneration

- 160.** Another alternative, though with less support in the consultation, is a broad exception along the lines of option 2 accompanied by a statutory licence or levy. Statutory licensing is an idea currently being explored by the Government of India. It would mean that AI developers would be able to use any lawfully accessible works to train their models – for any purpose and without right holder opt-out – but would be required to pay remuneration to right holders.
- 161.** However, many right holders and AI developers prefer to enter into exclusive licences, and it is likely to be challenging to identify a level of remuneration that is considered fair by all sides.

Focused exception: potential effects

- 162.** Below we consider the potential effects of a focused exception with lawful access requirement. As noted above, were a focused exception such as this to be taken forward, its scope and parameters would need to be clearly determined, and further evidence would need to be gathered in order to assess its impacts on right holders, developers and users, including individuals and SMEs. As such, the below description is only illustrative.

Data acquisition

- 163.** Under a focused exception, it could become easier to assemble certain training datasets, by allowing analysis of online content, but only for the targeted purpose, such as science and research. Lawful access provisions would mean right holders could control which works are available for use under the exception and which are not. As described above, additional safeguards could be included to reduce impacts on right holders.
- 164.** As a consequence, we would not expect a significant impact on the market for licensed datasets, especially those used to generate “lookalike” outputs. Overall, we would expect reduced effects on licensing income for copyright owners in the UK, including SMEs and individuals, compared to other exception options.

Pre-processing and training

- 165.** A focused exception could permit the making of copies to train or fine-tune a task-specific model. However, if focused on research, it could not be relied on for the training of general-purpose models, or fine-tuned models for a non-research purpose.
- 166.** Under this approach the training of general-purpose foundation models is likely to still take place in the USA, to the extent that USA law is considered more permissive. Many AI developers argue that a wide range of AI training activity is permitted under fair use, but case law is still developing. The scope of fair use may affect whether model training takes place in the UK under a focused approach.

Focused training and development

- 167.** The main beneficiaries of this approach would be people in the UK who develop and use AI for the specific tasks covered by the exception. Based on consultation feedback and our current understanding of the market, we expect beneficiaries are more likely to be UK-based SMEs, and individual developers and users in the UK, than the large firms that develop general-purpose foundation models.
- a) Fine tuning
- 168.** This approach could support certain applications of AI that take place in the UK, such as UK-based research. An example could be where a pre-trained AI foundation model is fine-tuned on a separate corpus of research data in the UK, which may lead to new scientific discoveries and new commercial applications.
- b) Data analysis and RAG
- 169.** This approach could also permit analysis by an AI tool, including RAG, where this was for the purposes covered by the exception. For example, it could be used by scientific researchers to locate and analyse information in papers they have lawfully accessed, and to ensure accurate citations.
- 170.** However, if the scope of any exception did not extend to general-purpose or competing uses, it is unlikely to extend to summaries of online content, including news and similar content. Access to such content could also be restricted using technical means. This content could still be used under licence, for example where websites opt-in to web crawling. Good practice on technical standards could give greater clarity and help to ensure those sharing or accessing content online do not face unnecessary uncertainty or risk.

Alternative approaches: assessment

- 171.** We acknowledge the views of those favouring an exception focused on the use of information contained in copyright works to advance innovation, knowledge and discovery, and other public interest outcomes.
- 172.** We also recognise other potential innovative approaches to copyright and AI, and the need to consider their effects.

- 173.** We will gather further evidence on alternative approaches. When considering any alternative approaches, we will consider their effects on right holders, AI developers and users of AI systems, including on individuals and SMEs.

Conclusions and Proposals

- 174.** Our approach to copyright and AI must enable the transformational benefits of AI, which will support growth and improve living standards, while protecting human creativity and our world-leading creative industries. We will not introduce reforms that do not support this objective.
- 175.** We recognise the strong opposition to the opt-out exception – the government’s preferred option at consultation stage – from many creators and the creative industries. There remains considerable uncertainty about the potential outcomes of this approach, and the ability of technological measures to make it work.
- 176.** In light of the strong views from the consultation, the gaps in evidence and the rapidly evolving AI sector and international context, a broad copyright exception with opt-out is no longer the government’s preferred way forward. Instead, we propose to gather more evidence on how copyright laws are impacting the development and deployment of AI across the economy and the economic benefits of reform. This will include further consideration of the effects of proposals on copyright owners, developers, and users, including the effects on SMEs and individuals.
- 177.** We propose to consider whether specific interventions are needed in light of that evidence and wider developments in the market here and abroad, including how these affect systems developed outside the UK and used within the UK. We will continue to keep them under review, in light of legal, technology, market and licensing developments in the EU, USA, and other countries.
- 178.** We propose to give further consideration to alternative approaches that have potential to support the government’s objectives.

Section D: Input transparency

1. This section discusses transparency in relation to the data and content used in the development and deployment of AI systems – i.e. the inputs. It considers the disclosure of information by AI developers, including about how they access and use copyright works.
2. The section considers potential approaches to these issues, including those taken in other countries. It assesses these approaches, taking into account views submitted to the consultation, and makes proposals for a way forward. Potential economic impacts of transparency interventions on right holders, AI developers, and AI users are considered further in the accompanying impact assessment.
3. In the consultation, input transparency measures were discussed as part of option 3 (the opt-out exception) and would also form part of option 1 (stronger copyright). Developers may also choose to introduce transparency measures under other options. The consultation included a number of specific questions on input transparency (questions 17 to 23). Many respondents also highlighted transparency in their comments on their preferred approach to copyright and AI (questions 1 to 5). Transparency is also a theme of the government’s technical working groups.
4. Transparency and AI-generated outputs, including the labelling of AI-generated content, are considered in the next section of the report. Technical tools and standards which may be used to support transparency are also discussed in more detail later in this report.

Input transparency

5. As described in the previous section, modern AI models are developed using large volumes of data, often amounting to billions of data points. Generative AI models will often be trained using creative works protected by copyright, such as images, text and music. Many AI developers use content which has been made available online, but which has not been licensed for AI training because they operate in jurisdictions where this is permitted.
6. There is often limited disclosure by developers about the sources of works used to train their models, and creators will often not know whether their works form part of a training dataset. This has raised concerns among right holders who believe their works may have been used to train AI models in breach of copyright law in the jurisdiction in which the model was trained.
7. It can also be difficult to understand which works have been used in other stages of AI system development and deployment. For example, online material and other datasets may be referenced during “retrieval-augmented generation” (RAG) by AI agents and internet search summaries. Right holders will not always be aware that their works have been accessed, or the process through which they have been collected.
8. Increased transparency about the use of copyright-protected works to develop AI systems was the main measure supported by right holders in the consultation. Greater clarity over which works are used, and for what purpose, might help them to enforce their rights if these have been infringed by the developer. This can include

information about the works that have been used, as well as how they have been collected, for example by web crawlers. Such information may be obtained through disclosure as part of litigation, but this can have significant costs for both right holders and developers, particularly small- and medium sized enterprises (SMEs)²⁷ or individuals.

9. AI model training and development often takes place in different countries. The current lack of consistent rules across jurisdictions on transparency is also a concern for others in the value chain. AI developers and dataset providers may feel that they will be disadvantaged if they disclose information and competitors do not. AI users, including SME and individual users, also have an interest in ensuring that a model they use has been produced in accordance with the law, and in understanding the provenance of training data.

Web crawler transparency

10. A web crawler is a computer program which is able to autonomously browse the internet and collect information from websites. A web crawler will systematically visit web pages, download their content and store it for various purposes, such as building search indexes for search engines.
11. Web crawlers have been used by AI developers, and providers of AI training datasets, to “scrape” large volumes of content from websites with which to train AI models. A significant amount of content scraped from websites will be protected by copyright and related rights.
12. Information on the nature and purpose of these web crawlers can help right holders to understand whether their works have been used to train AI models, helping creators exercise their rights. It can also help them to control access to their material, by blocking or allowing specific web crawlers.
13. Most major AI companies publish documentation about the web crawlers they use.²⁸ This helps website owners to control which crawlers can access their works and identify where their content has been accessed by AI developers. However, not all web crawlers follow this practice, or do so to different standards. Technical standards which enable website owners to control access are described later in this report.
14. AI developers are not legally required to disclose this information in the UK, and the level of detail made publicly available varies between individual companies.

Legislative approaches to transparency

15. The Copyright, Designs and Patents Act 1988 (The “CDPA”) contains provisions which aim to support the disclosure of certain information about copyright works. For example, authors, directors, and performers have moral rights to be identified as

²⁷ In this report, “SME” refers to micro (with fewer than 10 staff), small (at least 10 but fewer than 50 staff), and medium-sized (at least 50 but fewer than 250 staff) businesses.

²⁸ For example, details are published about OpenAI’s GPTBot, Anthropic’s ClaudeBot and Google’s GoogleBot and GeminiBot, among others.

such in certain circumstances (sections 77 and 205C CDPA) in relation to their works/films/performances, and certain exceptions, such as quotation fair dealing, apply on condition that the identity of the author of a copyright work is acknowledged. Deliberate removal or alteration, without authority, of electronic rights-management information such as authorship and terms of use of the work, in order to facilitate copyright infringement, is also prohibited (section 296ZG CDPA).

16. At present, there is no requirement in UK law for an AI developer or provider to publicly make available information about the copyright works used to develop an AI model, or how they were obtained. The UK's Competition and Markets Authority is currently consulting on potential requirements²⁹ which include an obligation on Google to be more transparent on how publisher content feeds into AI-generated search results, but such requirements would not apply more generally.

International approaches

17. The following section provides an overview of the approaches to input transparency which have been taken in the European Union (EU) and the State of California. As well as affecting their domestic markets, these approaches are likely to affect UK right holders and AI developers, as they apply on a “market access” basis. For example, a UK provider of a general-purpose AI system will need to comply with Regulation (EU) 2024/1689 (“EU AI Act”)³⁰ if they want to access the EU market. UK right holders and consumers will also benefit from disclosure requirements in other countries, aiding licensing, enforcement, and wider awareness of training data in the UK.
18. If the UK were to take a similar approach to regulation, then it could affect models developed outside the UK as well as those developed within the UK. There may be different effects on models developed in different territories depending on the types of models covered by the regulation. For example, regulations targeting foundation models are likely to have more impact on models developed in the USA, which is the source of many foundation models.

The European Union's AI Act

19. The EU was the first jurisdiction to adopt a comprehensive legal framework for AI transparency, through its EU AI Act. The EU AI Act aims to foster responsible artificial intelligence development and deployment in the EU, through a range of legal duties on different actors in the AI ecosystem.
20. Many of these duties fall on providers of “general-purpose” AI models – a term largely corresponding to “foundation” models described in Section B. Among other things, such models are capable of generating outputs such as text, audio, images and video, which resemble the content that they were trained on. The EU AI Act provisions relating to general-purpose models entered into force in August 2025.

²⁹ Competition and Markets Authority consultation: <https://competitionandmarkets.blog.gov.uk/2026/01/28/improving-the-way-google-delivers-search-services-in-the-uk/>

³⁰ The European Union AI Act: Regulation - EU - 2024/1689 - EN - EUR-Lex

21. The EU AI Act requires providers to put in place a policy to comply with EU copyright law, and requires transparency over various aspects of AI systems, including inputs, outputs, and the models themselves.
22. The European Commission has also published a Code of Practice, which aims to help AI providers to meet their obligations under the Act. The Code, which was prepared by independent experts in consultation with stakeholders, has a chapter aimed at meeting the copyright disclosure obligations of the Act.

Training data and web crawler transparency

23. Article 53(1)(d) of the EU AI Act requires providers of general-purpose AI models to draw up and publish “sufficiently detailed” summaries of their training material. Such a summary does not require a work-by-work assessment, but should be “generally comprehensive” and facilitate the enforcement of copyright and other rights.
24. The AI Office (a body of the European Commission) published a template to support this requirement on 24 July 2025. The template requires providers of general-purpose models to disclose:
 - The modalities of their training data (text, images, video, etc.), the size of each, the type of material covered, relevant languages, and the date of acquisition.
 - The main publicly available datasets (made available by a third party for free, not necessarily in compliance with copyright) used in training. For any large dataset this should include a link or other identifier, and a description.
 - Any web crawlers used to obtain content, including identifiers, and information about those crawlers including their purpose and the type of content obtained by them.
 - Other information, including whether data was collected from the users of a service, whether synthetic data was used to train the model, a description of rights-reservation methods, and a description of measures taken to remove illegal content.
25. AI providers are not required to provide information that would compromise their trade secrets.

California AI training data transparency

26. California’s Generative Artificial Intelligence: Training Data Transparency Act entered into force in January 2026. It requires generative AI developers to state publicly on their website a “high-level summary of datasets” used in the development of their system or service, including information on:
 - the sources and owners of the datasets;
 - whether the datasets include data protected by copyright, trade mark, or patent, or are entirely in the public domain;
 - other information, including whether datasets were purchased or licensed by the developer.

27. The Act does not include specific transparency obligations relating to web crawlers.

Views from the consultation

Input transparency

28. The government's consultation on copyright and AI sought views on whether measures should be introduced in the UK to ensure greater transparency over what content has been used to train AI models. It asked whether AI developers should disclose the sources of their training material, and what transparency should be required in relation to web crawlers. It asked how granular any information should be, and what a proportionate approach to transparency would look like. It also asked about the costs of transparency, and how it should be encouraged or enforced, and sought views on the EU's approach.
29. Most respondents - over 90% - agreed (in response to question 17) that AI developers should disclose the sources of their training material. This included a broad spectrum of the creative industries. The reasons cited varied and included determining the use of works (including respect for rights reservations), supporting enforcement of copyright, facilitating licensing, building public trust, and revealing model biases. Some of these respondents suggested that present non-disclosure of sources conceals copyright infringement.
30. Respondents from the creative industries generally took the view that transparency requirements should not be contingent on the introduction of a new copyright exception and should apply alongside existing copyright law.

Granularity and proportionality of disclosure

31. Respondents had differing positions as to the granularity and proportionality of potential disclosure requirements. A high proportion supported disclosure of information across the different stages of AI model development and deployment (including pre-training, fine tuning, RAG), including detailed information about the sources of training data. These views were not shared by AI and other technology firms, which generally supported higher-level and voluntary disclosure.
32. Those supporting information disclosure often said this should include information on datasets and websites that training data had been obtained from. Many from the creative industries also sought disclosure of individual works used by models, or information that would enable their identification, by providing a "line of sight" to them. For example, the News Media Association said that information should include digital object identifiers (DOIs) and file names, as well as "external metadata, proprietary metadata, and the name of the publisher it was scraped from". Many also highlighted the importance of recordkeeping to support disclosure.
33. There were also proposals about how information should be disclosed. Many respondents proposed a requirement on AI developers to provide a publicly available list of content they have ingested. They argued this would encourage developers to respect copyright, including rights reservations, and would also benefit downstream users of an AI system. For example, RELX, a publisher and provider of data analytics tools, said this "would not only provide transparency to

rightsholders; it would also allow any downstream users of the system to have full insight into the quality and authenticity of the works used in development”.

34. Some proposed a two-tiered approach to granularity: one level of public disclosure that outlines sources at database level and a more detailed level of disclosure about individual works to be provided in the event of a right holder looking to check for a specific breach. There were proposals that the latter could be provided through a method similar to a data protection “subject access request”.
35. Those who responded from the AI and technology sectors generally agreed that there should be greater transparency. But they advocated for higher-level and industry-led transparency and cautioned against burdens that would limit AI innovation or jeopardise security. TechUK said, “AI models vary significantly in function, and any regulatory framework must reflect this diversity with proportionate and appropriately scoped requirements, to avoid placing unnecessary burdens or undermining the security and integrity of AI applications”.
36. AI developers large and small cautioned that work-by-work disclosure would be challenging and costly to implement, particularly by SMEs or individual developers. Reasons raised by developers included the complexity of overlapping rights in items of content, where a single item may contain multiple works with multiple right holders, a lack of reliable metadata or registration of works to enable their identification, and the nature of AI development, which means that not all works collected will be used. The Startup Coalition noted that, while attempts could be made to automate disclosure of individual works, it will be challenging to do this accurately, so would require manual intervention. They were concerned that startups, which are often small or micro businesses, “do not have this luxury”, and new entrants would leave the market.
37. Open-source platforms, which make datasets available to the public for community-driven research and development projects, highlighted the benefits of transparency. These platforms typically require users to disclose documentation, including information on the content of the datasets and any licensing terms. Hugging Face said its transparency initiatives “demonstrate that proper documentation does not create undue burdens when appropriate tools are available. In fact, research on Hugging Face datasets has shown that well-documented datasets are also the most downloaded”. The Wikimedia Foundation also highlighted the benefits of transparency in relation to open datasets, including the benefits to AI users. This included attribution of sources in outputs, as a means to help ensure information is accurate and reliable and avoid bias.

Web crawlers

38. Most respondents supported transparency about the use of web crawlers. Right holders sought information that allows them to identify who has accessed their content, and for what purposes, and whether it was done in accordance with licensing terms and rights reservations. To support this, there were proposals for disclosure of the purpose, identity, and owners of any web crawlers, the timeframe or timestamp of web crawling, downstream users of crawled data, and the approach to rights reservations.

39. Some respondents noted practices such as “stealth crawling” (using unlabelled web crawlers) which allows for robots.txt instructions or other technical blocks to be circumvented.
40. Some internet infrastructure providers, such as Cloudflare, also called for greater transparency and accountability over web crawlers, including requiring AI companies to state their crawler’s purpose and provide contact information.

Other information

41. There were also proposals that AI developers share other information about their approach to copyright. For example, one respondent said that in addition to publishing reasonable information on the data used for training, “AI developers could also be required to maintain a policy regarding the measures they put in place to respect copyright while training.” Others sought information on other measures put in place by AI developers, such as steps to comply with technical standards.

Costs of disclosure

42. The prevailing view from the creative industries, and not shared by the AI sector, was that costs to developers of complying with transparency measures would be low. This included many statements that AI developers would already have the required information. For example, “AI companies should know the training data they have used, so that they can accurately pursue new data sources... It would merely be keeping a record of what they are doing already!”
43. Many technology industry responses were concerned with the practical and financial burdens of disclosure, particularly if detailed, work-level disclosure was required. They also flagged risks to trade secrets and confidentiality. This related to both compliance with and deployment of new tech.
44. For example, UK Interactive Entertainment (UKIE) said, “Any disclosure requirements must be measured, practical, and designed to balance transparency with feasibility. Excessive requirements could impose higher costs and disproportionately impact smaller AI developers. It is important that any requirement does not risk exposing trade secrets or violating confidentiality agreements”.

Compliance

45. There was significant support from the creative industries for legislative and regulatory underpinning of transparency requirements. Many felt that regulation should apply on a market-access basis, similar to the approach taken by the EU, which would mean that it would apply to models and services regardless of whether they had been developed within or outside the UK. Some voiced a preference for industry codes of practice or guidance with potential enforcement mechanisms for non-compliance. These included financial sanctions such as paying damages and regular auditing of AI companies.
46. For example, one respondent said “while voluntary guidelines may help, regulatory underpinning is essential to ensure AI companies do not ignore their obligations... Penalties for non-compliance, such as fines or restrictions on deploying models trained on unauthorized data, would provide a strong incentive for companies to follow the rules.”

47. More discussion of regulatory approaches to compliance can be found in the section on enforcement.

Views on the EU's approach

48. There were mixed opinions on the EU's approach to transparency. At the time of consultation, the EU was developing its Code of Practice on AI, and views were based on expectations of this process so may not reflect current opinion.
49. Many creative industry responses argued that the EU system was unlikely to be effective, although some acknowledged that it was a step in the right direction.
50. Several AI providers were of the view that close alignment to the EU could be a favourable outcome and, could assist with streamlining of developer obligations and increasing legal certainty. Others were concerned that having different transparency regimes in place in the UK and EU would be particularly detrimental to individual developers, micro, small and medium enterprises, who are likely to incur significant additional compliance costs.

Technical working groups

51. In addition to the stakeholder views gathered during the consultation, the government held four technical working groups with stakeholders during November and December 2025. One technical working group focused on transparency and facilitated a technical discussion with stakeholders on key issues including approaches to training data transparency and web crawler transparency. Another technical working group focused on the use of different technical standards.
52. In the transparency technical working group participants generally acknowledged the need for transparency to support licensing and enforcement. There was also general agreement that the transparency should be meaningful and actionable, although there was no consensus on the granularity needed to achieve this. Some participants, particularly those from the creative industries, suggested detailed records should be kept for right holders to access as needed, whereas others felt this was too burdensome, particularly for SME AI developers. Some expressed the view that transparency measures should not impede non-commercial research.
53. Some participants said that government regulation was needed for transparency to be effective. The group discussed the effectiveness of the EU's transparency regime, acknowledging that it is still bedding in, with some highlighting concerns and uncertainty on the approach taken. Views differed on whether transparency should be linked to copyright.
54. On web crawler transparency, right holder representatives explained the need for a regime that allows them to distinguish between different types of crawler, so they can have control based on the crawler's purpose.

Conclusions and proposals

55. Information about how AI developers train their models, including the content and data they use, can help developers demonstrate their compliance with copyright law and help right holders assert their rights. Currently, developers take a varied

approach to transparency, though requirements in other countries are beginning to support greater consistency.

56. In the consultation, the creative industries strongly supported the introduction of mandatory standards on transparency, who saw it as key to enforcing their rights. While technology companies also supported transparency, they argued that commitments should be high-level and industry-led. Views were mixed on whether we should follow the transparency approaches adopted by other jurisdictions, such as the EU. We propose to continue monitoring the effects of transparency rules in other countries and keep under review what, if any, changes to the UK's approach may be appropriate.
57. As part of our work to help right holders control and license their work, we propose to work with a range of industry and other experts to develop best practice on input transparency to help right holders assert their rights. Our approach should be proportionate and avoid unreasonable burdens, particularly on SMEs and individual developers, and promote clarity and enforcement for right holders of all sizes, including SMEs and individuals, without disincentivising AI development or deployment in the UK. The process should include developers of AI systems developed outside the UK.

Section E: Output transparency

1. This section discusses transparency in relation to the outputs of AI systems and copyright, including labelling of AI-generated content.
2. The section considers potential approaches to these issues, including those taken in other countries. It assesses these approaches and makes proposals for a way forward.
3. We draw on responses to the government's consultation. There were two specific questions on AI output labelling (questions 40 to 42).
4. Some of the measures described in this section may be supported through the use of technical tools and standards. This section focuses on the overall approach to transparency. Relevant technical tools and standards are described in more detail in Section F.

Overview

5. As the quantity and sophistication of AI-generated content increases, distinguishing it from human-created works has become increasingly difficult. In many cases it can be helpful for consumers to be able identify whether a piece of content has been generated using AI. Content labelling and other measures can help increase awareness of the type of content people are engaging with.
6. AI-generated content can be labelled by embedding information into it to show its provenance. This may be a visible label, such as a watermark with a company logo, or an invisible label, such as embedded metadata or a digital fingerprint. Such labels can assist in promoting consumer choice and help avoid misrepresentation. However, there are challenges relating to labelling, including the extent to which labels should apply to different media and uses, how to ensure labels are used, and how to ensure they are not modified or removed without consent from the right holders.
7. It may also not be clear which sources have been used to generate a specific output. While some services such as AI search may list sources of information, this will not always be consistent or prominent. Some right holders advocate greater transparency about the sources used to produce a specific output, though the feasibility of this will depend on the type of AI service.
8. The UK does not currently regulate how AI-generated content is labelled. But some developers and service providers provide labelling services and tools in the UK, either voluntarily or to comply with regulations in other jurisdictions. For example, Meta, TikTok, and Google have their own protocols for labelling AI content. The UK's Competition and Markets Authority is also consulting on potential requirements³¹ which include an obligation on Google to attribute the sources used to generate AI search results.

31 Competition and Markets Authority consultation: <https://competitionandmarkets.blog.gov.uk/2026/01/28/improving-the-way-google-delivers-search-services-in-the-uk/>

International approaches

9. Some countries have introduced legislation in this area, a selection of which is considered below. To the extent that foreign AI providers employ new measures in response to regulation in other countries, there may be spillover benefits to consumers and right holders in the UK. Likewise, UK developers will need to comply with regulation on output transparency in other countries in order to access those markets.

The European Union's AI Act

10. The European Union (EU) was the first jurisdiction to adopt a comprehensive legal framework for AI transparency, through Regulation (EU) 2024/1689 ("EU AI Act") that came into force in August 2024. The EU AI Act³² aims to foster responsible AI development and deployment in the EU, through a range of legal duties on different actors in the AI ecosystem.
11. Many of these duties fall on providers of "general-purpose" AI models – a term largely corresponding to "foundation" models described in Section B. Such models are capable of producing outputs which resemble the content that they were trained on.
12. The EU AI Act requires providers to put in place a policy to comply with EU copyright law, and requires transparency over various aspects of AI systems, including inputs, outputs, and the models themselves.
13. The European Commission Code of Practice has a chapter aimed at meeting the copyright disclosure obligations of the Act.
14. Article 50(2) of the EU AI Act places an obligation on providers of AI systems that generate synthetic audio, image, video or text content to ensure that the outputs of the AI system are marked in a machine-readable format and are detectable as artificially generated or manipulated. It excludes, for example, AI that provides an assistive function for standard editing, or that is authorised by law to detect, prevent, investigate or prosecute criminal offences.
15. Article 50(4) requires deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake, or text which aims to inform the public on matters of public interest, to disclose that the content/text has been artificially generated or manipulated. Where, in the case of image, audio or video content, the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, transparency obligations are limited to avoid hampering the display or enjoyment of the work.
16. These obligations do not apply where the AI use is authorised by law to detect, prevent, investigate or prosecute criminal offences or, in the case of text on matters of public interest, where the AI-generated text has undergone a process of human review or editorial control and someone holds editorial responsibility for its publication.

32 The European Union AI Act: Regulation - EU - 2024/1689 - EN - EUR-Lex

17. Work is ongoing to develop a Code of Practice on marking and labelling of AI-generated content to complement the EU AI Act.

California AI transparency legislation

18. The California AI Transparency Act requires certain AI providers to identify AI-generated content through visible or metadata-based labels where it is possible and provide free tools to users of their software enabling them to label and detect AI content. This obligation applies to any generative AI system accessible within the state that has over one million monthly visitors and applies to any content created or modified by the provider's system.
19. The Act requires disclosure in permanent or "extraordinarily difficult to remove" form (to the extent technologically feasible) of information including the fact the content was AI-generated, the name of the system provider, the name of the system, and date of creation.
20. Further legislation has been passed in California and other USA states requiring transparency of various other aspects of AI systems in order to support safety, accountability, and responsible AI development.

China's Deep Synthesis Provisions

21. China's Provisions on the Administration of Deep Synthesis in Internet-Based Information Services³³ ("Deep Synthesis Provisions"), came into effect in January 2023. Supporting measures and national technical standards on AI-generated content labelling took effect in September 2025.
22. They focus heavily on security and preventing the spread of misinformation. Deep synthesis is defined as the use of deep learning, augmented reality, and other algorithms to enable content synthesis or generation, encompassing text, images, audio, video, virtual scenes and more.
23. The Deep Synthesis Provisions apply to platforms, providers and users of generative AI. They impose labelling requirements on AI-generated outputs. These must be prominently displayed, to "remind the public of the deep synthesis situation." More stringent labelling requirements apply where the AI-generated outputs might "lead to public confusion or misidentification". Tampering with any labelling is prohibited.
24. Service providers are required to add explicit identification, clearly perceivable to users, and appropriate to the content type, to generated content. They are also required to add implicit labelling embedded into the digital content metadata, including information on the name or code of the service provider. Service providers are encouraged to use digital signatures within the metadata. They are encouraged to do this by adding watermarks to the generated synthetic content.

³³ Regulations on the Administration of Deep Synthesis of Internet Information: https://www.gov.cn/zhengce/zhengceku/2022-12/12/content_5731431.htm

Labelling of AI-generated content in South Korea

25. Korea's Framework Act on the Development of Artificial Intelligence, the AI Basic Act, came into effect in January 2026. It creates a two-tiered system based on the type of AI: high impact AI systems, which have safety and other public impacts, and generative AI systems.
26. It requires AI service providers to disclose to users, in advance, if their content is wholly or partly generated. These can take the form of notices or other technical measures.
27. Where AI systems generate virtual audio, images, or video that are difficult to distinguish from real content, AI service providers must notify or indicate to users, in a manner that allows clear recognition, that such content has been generated by an AI system. Where such outputs constitute, or form part of, artistic or creative works, such notification or indication may be made in a manner that does not interfere with their exhibition or enjoyment.

Views from the consultation

Degree of output transparency

28. Of the respondents that answered the consultation questions on output labelling, the vast majority supported the labelling of wholly generative AI outputs. Reasons included reducing the risks of deepfakes, misinformation and other potential harms, building user trust and encouraging responsible use of AI, and enabling the public to make informed decisions on how to interpret and use information.
29. There were more differences in opinion when it came to AI-assisted works, where AI is used as a tool to assist human creation. A few respondents suggested labelling requirements should be in place for all content that includes an element of generative AI in a final work. But a number suggested there should be a distinction between wholly AI-generated content and AI edited or assisted content, which should have more flexible and optional labelling requirements.
30. For example, Adobe said that "wholly AI-generated content warrants automatic labelling" whereas providers of AI-editing services "should be required to provide the user with AI labelling tools and allow users to attach a label". They noted that automatically attaching a label to any type of edited content could cause confusion and devalue a creator's work by implying that the whole thing is AI-generated. They also supported leveraging industry standards such as Content Credentials. These views were shared by other providers of content-creation software.
31. Many creators and groups representing them drew similar distinctions. The Association of Photographers said, "It is important to ensure that labelling is only applied to machine-generated outputs and not where human-authored works, such as photographs, which have AI-assistive tools applied in order to edit photographs, as these copyright-protected works are expressive works made by humans." They warned that certain tools and contract terms do not allow for such distinctions, which can disadvantage creators who use AI tools to edit their works by giving a misleading impression that such a work is AI-generated.

32. Other respondents suggested labelling should be limited and only be required where there is the risk of public harms and not in other areas where consumers know the content is fictional and has AI driven mechanics. Sectoral examples were given such as the use of AI for CGI in film and television, and for video game graphics. For example, Pact, representing film and TV producers, said AI labelling should only be required “where there is a tangible risk of harm (i.e. to combat deepfakes or ‘passing off’) ... For example, the production sector does make use of AI-assisted tools to edit background colours or alter the appearance of actors. Labelling such content as AI generated would not be practical within the film and TV sector.”
33. A number of respondents commented on limitations to labelling solutions in relation to different types of content. For example, OpenAI said “Provenance solutions must be accurate, avoid degrading content quality, and resist tampering, and the right balance for these three objectives will vary by modality. We have developed solutions that seek to meet these goals for images, audio, and video, but we have not found a solution that works for text. ... we are not aware of labelling methods for text that are both accurate and cannot be circumvented by bad actors, such as by rewording text using another generative model”.
34. Others shared similar views, noting the need to consider the modality of content and the current state of the art in labelling technology. Many agreed that labelling measures should avoid degrading content and should be resistant to removal, to greatest extent that is possible.

Compliance

35. Most respondents, particularly from the creative industries, supported some form of regulation for AI output labelling and suggested the current lack of regulation has led to a fragmentation of different voluntary approaches. A number of respondents suggested regulation should ensure labels are automatically embedded as watermarks and into the metadata of AI-generated outputs by the AI system. Many stated this should include details of the AI tool used to generate the output. For example, the Creators Rights Alliance said “All work should be labelled to ensure transparency and allow audiences, consumers, business, and rights holders to distinguish between human-created and AI-generated works. A proportionate approach would require AI developers to implement automatic metadata tagging and watermarking at the point of creation.”
36. Others suggested legislation that is limited to specific platforms – for example where AI outputs are posted on social media or on news sites. Suggestions also recognised the need to work with industry and take into account technological development. Suggestions included setting minimum standards within principles-based frameworks, without prescribing specific technologies. Several respondents favoured co-designing voluntary good practice with industry, with a regulatory approach if such good practice was not adopted. For example, AI Governance Limited said “We’d like to see education on good practice to encourage appropriate labelling as well as pressure to develop industry standards and guidelines for labelling. If this did not work then regulation would be appropriate.”
37. Some respondents focused on labelling tools and standards. These are covered in more detail in Section F of this report. Many respondents expressed support for the C2PA Content Credentials initiative. Others mentioned the use of metadata

standards such as or the International Standard Name Identifier (ISNI). Some suggested government should have a role in supporting new technologies, noting that labelling was not the only solution. For example, the Oxford Intellectual Property Research Centre said “It is relevant to mention that this function of attribution will best be achieved if the use of watermarking technology is accompanied by other provenance-facilitating technologies. Therefore, we suggest that developing watermarking technologies ought to include research and development into other provenance technologies as well.”

Technical working groups

38. In addition to the stakeholder views gathered during the consultation, the government held four technical working groups with stakeholders during November and December 2025. Labelling of AI outputs was one of the issues discussed during these technical working groups.
39. Stakeholders acknowledged that identification and provenance tools are developing and will be key to improving output transparency. Most viewed the labelling of wholly AI-generated outputs as a helpful way of increasing awareness of the type of digital content consumed. Some stakeholders gave more nuanced positions when AI is used as an assistive tool to produce content, with some cautioning the risk of over-labelling the use of AI.

Conclusions and proposals

40. We welcome the positive engagement that we have had with stakeholders on the issue of output transparency, both through the consultation and through our technical working groups.
41. We propose to work with industry to explore good practice on labelling AI-generated content. The aim will be to identify balanced and effective practices that help to establish consumer confidence and public trust in the outputs of AI.
42. We propose to continue monitoring international developments to ensure our work is complementary to, and not duplicative of, other processes, which will help set global standards. We also propose to work with international partners where possible to support the development of common solutions.

Section F: Technical tools and standards

1. This section explores the role of technical measures and standards in controlling the access to, and use of, copyright works for the purpose of developing AI systems. In this section, we use the term “technical tools” to describe technical measures right holders and website owners may use to help them to control access to works, how they are used, and under what conditions. “Technical standards” are agreed guidelines and procedures that promote consistency, interoperability and trust, enabling technical tools to be used effectively. This section considers the potential role of such tools and standards if reforms are made to the copyright framework. It also considers good practice in this area and how it could be encouraged.
2. In the consultation, technical tools and standards were discussed as part of option 3 (the opt-out exception), but they may also have a role supporting other options, as discussed below. The consultation also included questions on technical standards (questions 9 to 11). This section draws on responses to these questions, as well as other sources. It also summarises the activity of the government’s working group on Control and Technical Standards. Potential economic impacts of measures relating to technical tools and standards are considered further in the accompanying impact assessment.
3. Tools and standards continue to develop rapidly. However, current tools and standards do not support all right holders’ needs, and there are challenges with adoption and compliance. We will continue to consider the need for regulation in this area and work with the AI and creative sectors to support best practice and the adoption of technical tools and standards.

Overview

4. Much of the debate on copyright and AI relates to the use of creative content available on the internet and other public platforms. Right holders often have difficulty in controlling access to their content, and the use of their content in accordance with their wishes. Technical standards and measures or tools, including metadata associated with content, are increasingly being used by right holders to enforce their rights and express their usage preferences.
5. As discussed earlier in this report, many foundation models are trained using text, images and other content gathered from the internet, via the use of web crawlers. Web crawlers can be deployed by model developers themselves, or by organisations which compile datasets for AI developers. The use of tools and standards associated with digital works online interacts with AI development domestically and outside the UK.
6. Web crawlers are also used for retrieval of information for the purpose of factual “grounding” for “retrieval-augmented generation” (RAG) by AI agents and internet search summaries. This is where the internet is used as a data source in response to a specific request, and a pre-trained LLM uses the additional data to provide an up-to-date response.
7. Technical tools and standards are increasingly being used by right holders to manage access to, and use of, their works online. Their uses include:

- Reserving rights ('opting-out') of the EU's data mining exception under Article 4 of the Digital Single Market Directive³⁴. This exception requires right holders who wish to reserve their rights to express this in a "machine-readable form". Similar requirements could apply were such an approach to be adopted in the UK (see discussion of consultation option 3, in Section C of this report).
 - Helping to enforce copyright by preventing access to, and use of, works against the wishes of right holders.
 - Signalling permitted uses and licensing terms for copyright works. This can help right holders to allow some uses, to block others, and to allow others upon payment.
 - Providing information on the provenance of works, such as the author and nature of any use of AI.
8. Technical standards are collectively agreed procedures which promote consistency, interoperability, and trust, enabling technical tools to be used effectively. Some standards are approved and developed by organisations such as the International Organization for Standardization (ISO), and the Internet Engineering Task Force (IETF). Technical tools describe technical measures right holders and website owners may use to help them to control access to works, how they are used, and under what conditions and often implement or interact with technical standards, helping them to function reliably and predicably.
9. The following section considers types of technical tools and standards that are relevant to both right holders and AI developers. We have grouped them into three broad approaches: site-based, unit-based, and registered or notified.

Site-based standards

10. Site-based, or location-based, standards and tools enable website owners to control access to, and use of, content at the website or domain level. Site based permissions apply across an entire website or parts of a website.
11. An example is the Robots Exclusion Protocol (REP). This allows a website owner to control web crawlers' access to content hosted on their website. It is implemented through a text file called robots.txt, which is placed at the root of a website domain. A web crawler can read the contents of a robots.txt file and follow its instructions on which crawlers are allowed to crawl a site, and which parts of a site they are allowed to access. They may also include other details such as how many requests a crawler can make.
12. The REP is a voluntary standard but has been widely adopted by industry. Robots.txt files can be viewed by adding "robots.txt" to a web domain (for example, www.gov.uk/robots.txt). Increasingly, commercial content providers such as news websites are employing more detailed and sophisticated robots.txt files, which

34 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

disallow web crawlers used for AI training and RAG search summaries (for example, <https://www.ft.com/robots.txt>).

13. Robots.txt files are increasingly being used to disallow AI-focused web crawlers from crawling websites. This is being done directly by right holders who control their own websites, such as news media sites, and by hosting platforms. Several popular platforms enable their users to block common web crawlers used for AI training and RAG through simple means, such as a tick box.³⁵
14. Sites may also employ additional tools and methods to block crawlers which do not respect the REP. For example, Cloudflare offers a service which they launched in July 2025 to enable website owners to control how web crawlers access their websites. The default approach blocks AI web crawlers from websites that it hosts. However, website owners will be able to allow or disallow crawling for each stage in the AI life cycle such as training, fine tuning and inference. There is functionality to enable specific, verified web crawlers to train on their website content.
15. The IETF, an international standards organisation, has established a working group to develop a set of standards that would allow more granular control over different types of web crawling activity. One part of this work would result in a revised version of the REP. The aim is to agree standards that allow preferences to be set for different purposes for example, for web crawlers used for traditional online search and those used to collect data to train and fine-tune AI models.
16. This approach would mean a website provider would not need to block individual crawlers, and keep lists of crawlers up to date, but could block or permit categories of crawlers according to their preferences. However, at the time of writing the group had not reached consensus on certain key issues such as how to treat RAG.
17. Additional standards can be built upon robots.txt. A recent standard which builds upon the robots.txt framework is the Really Simple Licensing standard, which can be added to websites to specify licensing terms for content. Such terms can indicate the purpose that the site content used for, as well as the conditions, such as acknowledgement, payment, etc.

User-agent: Examplebot	# all Example services
Disallow: /private/	# disallow this directory
User-agent: Examplebot - news	# only the Example news service
Disallow: /	# disallow everything
User-agent: *	# any robot
Disallow: /this/	# disallow this directory

Figure 3 - Example of a robots.txt file addressing multiple user-agents

³⁵ E.g. Squarespace - <https://support.squarespace.com/hc/en-us/articles/360022347072-Request-that-AI-models-exclude-your-site>

Unit-based standards

18. Unit-based standards offer right holders and creators more granular control over their works, by associating permissions or restrictions directly with individual items of digital content such as specific images, songs and text files.
19. An example is where machine-readable metadata is embedded in the header of a media file. The metadata will be readable wherever the file is found, not only on a specific website.
20. Such metadata may describe details such as authorship and ownership, and increasingly, usage conditions and licensing information for a specific work.
21. In an AI training context, metadata can include reservations on the use of the work – for example, to flag to an AI model developer that the specific work is not available for AI training, or that AI training is permissible by obtaining a relevant licence. Various standards for different types of content including text, images, video and music have been developed.
22. Examples include the industry-led TDM Rep Protocol specification that has been developed in response to the EU's TDM machine readable rights reservation legislative requirements. It is a site- and unit-based specification. It provides a means of declaring whether TDM activities can be carried out via different techniques including embedded in EPUB files, image and HTML pages with potentially different permissions for specific digital content. For example, an HTML page may have different permissions to an image embedded on that page.
23. In the music industry, DDEX³⁶ is investigating metadata requirements for the communication of AI generated music. This information, which is associated with individual recordings and musical works, could enable right holders and those involved in music production to communicate AI-generated music across the music industry digital value chain. The Intellectual Property Office is supporting industry efforts to improve metadata in music streaming³⁷.
24. Tools that provide additional protection for artistic image works include Glaze and Nightshade³⁸ which allow creators to subtly change their images to provide protection against style extraction and alter AI training signals.
25. However, there are challenges with unit-based approaches. Metadata may be stripped from a digital file, which will remove creator and right holder information and preferences. Adherence to metadata, and its onward provision, relies on implementation by platforms and intermediaries. Right holders and others may lack the technical knowledge to apply unit-based controls to works and at present such measures are voluntary and not consistently adopted.

Registered or notified data

36 Artificial Intelligence Ad hoc Group (AI) <https://ddex.net/about-ddex/working-groups/>

37 Artificial Intelligence Ad hoc Group (AI) <https://ddex.net/about-ddex/working-groups/>

38 TheGlazeProject <https://glaze.cs.uchicago.edu/aboutus.html>

26. Some AI firms and dataset owners also offer right holders the ability to notify them more directly that they do not want their works to be used for training AI. Sometimes this requires that right holders notify AI firms about individual works. But technologies are available which simplify the process of notification, by allowing notification to a central registry.
27. The European Commission is also exploring the potential for a centralised “opt-out” registry. It has commissioned a feasibility study for a registry that would enable right holders to express opt-outs at the work level and facilitate their identification by AI developers.
28. One advantage of this type of approach for right holders is that it can be used for block training on works which are already present in an AI training dataset, regardless of the metadata that was associated with them when they were obtained. However, such registers require greater infrastructure for their operation than standards-based approaches and face similar adoption challenges.

Current challenges

29. The use of online material by generative AI, including retrieval-augmented generation, has led increasing numbers of right holders and web hosts to adopt technical measures. It has also led to a more active marketplace for different standards and technical solutions.
30. However, there remain challenges to the development and implementation of different types of tools and standards, which affect their adoption and use by different groups.

Agreement of standards

31. To develop and finalise a standard there must be broad consensus among a range of stakeholders on its contents and definitions. In an area such as AI, which is rapidly evolving, it can be difficult to determine what should be in a standard and how to express rules and definitions. It can also be challenging to reach consensus where groups have different interests and incentives.

Adoption (by right holders and technology companies)

32. For an agreed standard to have its intended effect, it needs to be adopted by a sufficient number of players in a market. A new standard on use of content with AI needs to be adopted not only by content creators, but also intermediaries (including social media and other platforms), and AI developers. For example, a music standard may not be effective if it is only used by recording artists and labels but not by streaming platforms, which are the main digital distributors of music. As raised in a number of responses to the consultation, SME and individual right holders may find it more difficult to adopt new standards, as they may lack the time, resources and knowledge to do so. SME and individual developers may also find it more difficult to put in place measures that comply with standards and to keep pace with new standards developments.

Trust

- 33.** Most standards are applied and enforced voluntarily. Those using them want to have confidence that others are using them appropriately. A content creator who is applying metadata to their work will want to know that this is worthwhile and will not be ignored. Digital platforms, developers, and consumers will want to know that any metadata that has been applied to a work is accurate. It is important to develop and foster trust between right holders, creators, AI developers, and users. Issues relating to trust between right holders that apply metadata and downstream users of their content were raised in the consultation. For example, some right holders noted that metadata is often removed or replaced when content is uploaded to social media or user-generated content platforms.

International approaches

- 34.** The development and use of technical standards in the field of copyright has usually been industry- and market-led. The government has supported industry-led efforts to improve the use of metadata.
- 35.** The European Union is also taking action to promote the development and use of technical standards in relation to AI and copyright. This is in the context of its opt-out copyright exception and the requirements of its AI Act.
- 36.** Commitments to good practice in relation to technical standards are part of the EU's AI Code of Practice. Measure 1.3 of the copyright chapter requires signatories to commit to ensuring web crawlers read and follow rights reservations expressed via the Robot Exclusion Protocol as specified by the IETF and any future versions of this protocol. It also includes requirements to identify and comply with unit-based metadata rights reservations. The European Commission recently launched a consultation on rights reservation protocols, to support implementation of the EU's AI Act and Code.³⁹
- 37.** The EU's AI Act and Code apply to providers of AI systems to the EU, regardless of whether they were developed in the EU or elsewhere. As such, UK AI developers that deploy their models in the EU will be expected to comply with these provisions. Although the rules apply regardless of where a model is developed, there may be different impacts on models developed in different countries – for example, the general-purpose foundation models caught by the regulations are more likely to be developed in the USA, and models developed in the EU may be more likely to be deployed there. As they apply to general-purpose models, they are also more likely to affect the larger firms which develop these than SMEs and individuals developing specialist applications.
- 38.** Developers of AI systems in the UK and outside the UK are likely to employ the same technological measures for the services they offer across jurisdictions. Therefore, regulations in the EU and elsewhere are likely to provide some spillover benefits to UK right holders and consumers, of all sizes.

Consultation options

³⁹ EU's AI Act code: <https://digital-strategy.ec.europa.eu/en/consultations/commission-launches-consultation-protocols-reserving-rights-text-and-data-mining-under-ai-act-and>

39. The Copyright and AI consultation sought views on the use of technical tools and standards to support greater right holder control over the use of AI with copyright works. It asked whether there should be greater standardisation in this area and, if so, what action the government could take (consultation questions 9 to 11).
40. Technical tools and standards have a role in each of the consultation options. If the current law is maintained (options 0 and 1) we expect technical standards to continue to be developed by the market, to control and enforce the use of content. We expect them to become more widely adopted, though the extent is unclear. We can see this trend already with AI crawling increasing by fifteen times in 2025 and AI crawlers were the most frequently fully disallowed user agents found in robots.txt files from the 2025 Cloudflare Radar Year in Review.⁴⁰
41. Technical standards are crucial to the effective functioning of an exception with rights reservation (option 3), which requires opt-outs to be machine readable. Technical tools and standards also have a wider role under this option, to enforce usage conditions and licensing terms, for example.
42. Under other exception options, which do not have rights reservation, technical standards will also play an important role in controlling access to works. In Section C, we discuss alternative approaches including a potential targeted exception with a defined lawful access requirement. Under such an approach, adherence to technical standards good practice could be part of assessing whether works have been accessed lawfully and in good faith.

Views from consultation

Machine-readable opt-outs

43. The majority of consultation respondents expressed support for standardisation if an exception with rights reservation were implemented, because it would reduce complexity or ambiguity of opt outs. However, many expressed this view while rejecting the exception with rights reservation as their preferred option. Many respondents who criticised machine readable rights reservations did so because – in the context of an exception with rights reservation – they felt it would place an unreasonable administrative burden on right holders, who would have to apply technical restrictions to their works. Many of these respondents also expressed views on the use of technical measures to support enforcement and licensing of copyright works under the status quo.
44. Many creative industries responses stressed the importance of technical solutions that minimise administrative burdens. They were concerned that right holders might face significant costs in implementing any standard across multiple platforms and in monitoring whether such reservations were respected by AI developers. There were concerns that such costs would affect individual right holders and SME right holders in particular, who lack time, resources and technical capacity. Respondents advocated solutions such as central registers, and enabling preferences to apply to

⁴⁰ The 2025 Cloudflare Radar Year in Review: The rise of AI, post-quantum, and record-breaking DDoS attacks <https://blog.cloudflare.com/radar-2025-year-in-review/#ai-user-action-crawling-increased-by-over-15x-in-2025>

multiple works simultaneously. Some suggested exploring whether natural language statements such as ‘all rights reserved’ are, or could as technology improves, be machine readable.

45. The British Copyright Council also noted challenges in how standardisation is “applied to the multi-tier combination of copyright works” in composite content, such as sound recordings and films.
46. Some suggested that the administrative burden should be placed on AI developers instead of right holders. An example was the use of automatic content recognition (ACR) technology to establish whether content is protected by copyright, who the relevant right holders are, and how to obtain permission. The BPI, a UK record label trade association, suggested ACR has “multiple uses. It is already widely used by online platforms to identify and filter out user uploaded content that infringes copyright. Stable Audio Open AI model also used ACR tools to remove copyright-protected music from a dataset before training.”
47. Many right holders were sceptical that technical standards were the solution to creator control in the context of an opt-out exception but thought that compliance with them more broadly would bring benefits. For example, the News Media Association supported enforcement of technical standards (particularly web crawlers) to help ensure copyright compliance.
48. Multiple responses highlighted concerns about removal of ownership and other metadata. Respondents from the music sector noted that metadata associated with a recording of a song may be removed by a third party each time a copy is made, for example, when it is dubbed into a user generated content platform. The Association of Photographers noted that unit-based data attached to a digital file “is not robust, cannot be locked, and is editable at a later date”, and is frequently removed when uploaded to a platform.
49. Groups representing AI firms also raised concerns that rights reservation mechanisms will be costly to administer for smaller AI developers. The Startup Coalition noted that many startups (which are often micro, small or medium sized businesses) were concerned with the additional resources it would take to implement technical standards, and that it may be more difficult for them to comply with requirements than for larger firms.
50. Some respondents supported specific technical standards, both in the context of rights reservation and more widely. For example, RELX proposed the TDM Rep specification for its ability to signal opt-outs. They described it as “a simple and practical web protocol, capable of expressing the reservation of rights relating to text and data mining applied to lawfully accessible web content”, as well as easing discovery of licensing policies.
51. Several broadcasters and visual content platforms expressed support for C2PA Content Credentials. Adobe noted that Content Credentials could be used to “attach a “do not train” tag to an individual piece of content that remains associated with the work wherever it is published online”. Content Credentials are also becoming widely adopted as a means for labelling content and signalling authenticity (see Section E of this report).

52. Many AI firms also expressed their support for technical standards to implement an opt-out exception. They tended to support the Robots Exclusion Protocol (REP) as the primary means for indicating opt outs, as they argued other means were not yet sufficiently developed or widely adopted. The REP was the most commonly referenced protocol in consultation responses generally. For example, Meta said they ‘support robots.txt as a widely used and understood standard’. Google said “Overwhelmingly, the feedback we have received from site owners is that they understand robots.txt and its mechanisms, and are able to use it appropriately to achieve their goals”.
53. As an example of alternative tools to provide right holders with control of their content, Cloudflare highlighted their AI Audit tool which “blocks the crawlers at the network level through a web application firewall rule” and allows site owners to “see precisely how often their sites are being scraped, and by whom, they can decide which of the crawlers (if any) they want to enable access to”.
54. A range of other respondents also expressed their support for technical tools and standards to aid content control, transparency, and licensing, including organisations which provide content under open licences, such as Wikimedia.
55. Some respondents also commented on the wider impacts of increased use of technical standards on users of internet-based services. They cautioned that this may lead to currently accessible content becoming less accessible to users and leading to “walled gardens” where content is accessible only through subscriptions and other means. An outcome where less content is available on the internet for free could have disproportionate effect on individual users and SMEs.
56. Many of those supporting standards noted challenges with their adoption and enforcement, and recommended government intervention to support this.

Licensing

57. Technical tools can support signalling to AI developers of permitted uses and licensing terms for copyright works in machine readable formats. This can help right holders to allow some uses, to block others, and to license other uses in exchange for payment. The use of technical standards to promote licensing was encouraged by several consultation respondents.
58. For example, the Association of Learned and Professional Society Publishers favoured standards which “specify how to approach the right holder to obtain a licence”. RELX noted that this was a function of the TDM Rep specification. More recently (and subsequent to the consultation) a range of stakeholders have highlighted the RSL standard as a basis for content licensing.

Compliance with technical standards and protection of metadata

59. Many respondents expressed concerns that some AI developers’ activity seek to circumvent rights reservation and access restrictions. The News Media Association said some AI developers are “impersonating reputable AI bots, using techniques that

impersonate the human readers publishers are trying to reach, using headless browsers to evade identification, using proxies to get around geolocation and IP blocks, evading CAPTCHA security measures, and switching between IP addresses.”

60. Several responses suggested that technical standards and metadata that signal restrictions on the access to, and use of, content must be protected in law against circumvention or removal. The CDPA already protects “effective technological measures” and “rights management information” against circumvention and removal, with potential civil remedies and criminal penalties. Many respondents expressed a view that there should be more enforcement including penalties or sanctions against such removal. Representative groups such as the Alliance for IP and British Copyright Council said that systems for rights reservation should be given the same status in law, including criminal penalties when they are circumvented.

Potential government intervention

61. Several responses proposed regulatory approaches to enforcement of technical tools and standards. Many creative industry responses called for damages for non-compliance with technical signals such as rights reservations. There were also repeated calls for injunctive remedies including removal of works from training datasets or models if found in breach of them.
62. However, others expressed concerns about standardisation. For example, the Association of Illustrators voiced concern that standardisation could risk “potential technical monopolies that could increase costs for creators”.
63. Most AI developers argued that a non-legislative, standards-based compliance framework would allow AI firms to train their models without unnecessary legal risk. They cautioned fixing a specific standard in legislation, to allow flexibility and avoid obsolescence, pointing instead to industry led initiatives like the IETF extensions to robots.txt. The Centre for Decentralized Digital Economy noted that where legal requirements are “required by law, legal certainty and safe interoperability is enhanced, but potentially at the cost of flexibility and innovation”.
64. Some respondents suggested government could issue guidance that could include a list of appropriate standards identified in cooperation between industry and right holders. Adobe compared this approach to the FCA in their “guidance to financial services companies on how to comply with financial services legislation”. They said that any guidance should be “technology-neutral, globally harmonized and focus on the concepts and principles it aims to uphold”.

Control and technical standards working group

65. One of the government’s technical working groups was asked to consider Control and Technical Standards. The first meeting took place 25 November 2025, and was chaired by Professor Tom Crick, Chief Scientific Adviser at DCMS. These discussions were convened under the Chatham House Rule, which allows sharing of information without disclosing the identity or affiliation of the speakers.
66. In this discussion, there was some support for open standards and traceability of digital works. A number of attendees supported unit based technical tools due to

the greater granularity of control they can provide for digital works. However, concerns were noted with metadata stripping from digital works which could limit the effectiveness of these tools, especially in the downstream context in the AI ecosystem. Others supported domain level standards, and the IETF work on extensions to robots.txt was highlighted as an example of a positive development. Some concerns were raised on behalf of individual creators and right holders to ensure any tools can best support their needs, such as ease of use and potential cost to implement technical standards.

67. There was discussion on the potential role for government going forwards to support the development and implementation of technical standards. Some attendees suggested the government should influence and support interoperable standards internationally. Others cautioned that this needed to be balanced against the need to maintain space for industry and international standards organisations to continue to innovate tools and standards. There was some support for legislation on compliance to respect terms embedded with technical tools at both domain and unit level. Other attendees referenced the EU's AI Code of Practice and suggested the UK could adopt a similar legislative approach.

Conclusions and proposals

68. The government welcomes the pace of development of technical tools and standards that has taken place since it published its consultation. It continues to support the use of these tools to support right holders in enforcing their rights and managing permissions and licensing for AI training and development.
69. We recognise that the development of such tools and standards is primarily for the market and the government should take care not to stand in the way of the innovation that is taking place.
70. However, we consider that there remains a role for government to support industry to develop best practice on the use of technical tools and standards, as well as encouraging their adoption.
71. We propose to take forward work with experts and stakeholders to support best practice and adoption of market-led tools and standards. This work will seek to encourage the adoption of tools and standards that enhance creator and right holder control, facilitate licensing, support wider objectives on transparency and AI output labelling, and include developers of AI systems within and outside the UK.
72. Best practice could include principles that technical standards should comply with and identify standards that meet these principles. It could cover the use of technical tools and standards to control access to and use of works, including practice around the use of web crawlers. It could also cover wider issues relating to standards, such as their role in licensing content. Any best practice would need to take into account the needs of and likely effects on SMEs and individual developers and right holders and users, with the aim of avoiding the imposition of unreasonable or disproportionate burdens.
73. As with the proposals on transparency and labelling, we propose to keep the need for regulation in this area under review, and to continue to monitor the effectiveness of approaches taken in other countries.

Section G: Licensing

1. This section explores the role of licensing in accessing and using copyright works for AI training. It provides an overview of the requirements to license under copyright law and describes the current licensing environment with reference to established and emerging licensing models. Licensing and its impacts on different groups is also considered in the accompanying impact assessment.
2. Licensing is key in ensuring creators are paid when their work is used and plays an important role in incentivising new creative content. Licensing for AI is a new and evolving market that continues to grow, with opportunities for both right holders and AI developers.

The legal basis for copyright licensing

Statutory provisions

3. As described in Section B, copyright provides several exclusive rights to its owner, which enables them to control how their work is used. Where permission is required to use a work, it is typically granted by way of a licence.
4. In the UK, the first owner of copyright in a work is generally its author but it can be assigned or licensed, in whole or in part, to another party. This means that the primary creator is not always the right holder with authority to license the use of their work, including for AI training.
5. Most rights set out in the Copyright, Designs and Patents Act 1988 (CDPA) are exclusive rights. This means that the copyright owner has the right to authorise or prohibit the use of their works, and under what conditions. However, there are some areas where copyright law determines the conditions under which licensing should take place.
6. For example, when a sound recording is broadcast or played in public, a performer in that recording has a right to “equitable remuneration” from the sound recording owner (section 182D CDPA). This is not an exclusive right, so they cannot prevent use of their performance, but they are entitled to a payment for it.
7. Statutory schemes also exist to support the granting of licences for specific purposes. The government can authorise collective management organisations (CMOs) representing a significant number of right holders for a category of works to issue “extended collective licences” (ECLs), which cover all works of a certain type. The orphan works licensing scheme enables the government to grant licences under conditions where the right holder is unknown or cannot be traced.

Regulation of licensing

8. Licensing is a commercial activity which is typically market-led and without government intervention, unless market failures require it. The government has a limited regulatory function under the Collective Management of Copyright (EU Directive) Regulations 2016. These set light touch minimum standards for the governance, transparency, and conduct of CMOs, and other specified organisations. Collective licensing is discussed below.

9. There is also an independent judicial body, the Copyright Tribunal, which adjudicates on UK commercial licensing disputes between copyright owners or CMOs and people who use or want to use copyright material in their business.

International approaches

10. As described in Section B, copyright is territorial, meaning that it is usually the law of the country within which an act takes place that applies. There is a high degree of convergence of the rights which are available under national copyright laws due to several related international treaties. However, the treaties permit some flexibility in national implementation, including exceptions to rights in some circumstances.
11. These treaties also mean that the rights of UK right holders are generally protected on the same terms as countries' own nationals. This means a UK right holder is able to license their works in other countries, and will be affected by any exceptions, licensing schemes, or other measures that apply there.
12. The deployment of AI systems which have been developed outside the UK (where training would not have met the requirements of UK law) may give rise to copyright infringement under UK law if models imported into the UK contain or comprise infringing copies of copyright works (section 22 CDPA).
13. The recent judgment in *Getty Images vs Stability AI* confirms that, in principle, an AI model can constitute an "article" for the purposes of secondary infringement under section 22 CDPA. However, in finding that Stable Diffusion did not contain any infringing copies and therefore was not an infringing copy, the case highlighted novel questions about the extent to which AI models store copies of copyright works. The judgment is currently subject to appeal on the issue of what constitutes an infringing copy for the purposes of importation under UK law.
14. There is no statutory licensing scheme for the use of copyright works to train AI models in UK law, and we are not aware of any that exist in other countries. There is also no UK precedent for a statutory copyright levy. However, in late 2025, the Indian Government consulted on a statutory licensing proposal. If taken forward, this would make available in India all lawfully accessed copyright content for AI development, as a matter of right, with fair compensation for right holders to be determined by a government-appointed committee and paid post-commercialisation of the trained AI system. It would also impose transparency requirements on AI developers to disclose the datasets used for AI training.
15. The Australian Government is consulting a 'Copyright and AI Reference Group' on priority areas which include "encouraging fair, legal avenues for using copyright material in AI". Specifically, it is examining whether a new collective licensing framework should be established, or whether to maintain the status quo through a voluntary licensing framework.^{41 42}

41 Australian Government announcement: <https://ministers.ag.gov.au/media-centre/albanese-government-ensure-australia-prepared-future-copyright-challenges-emerging-ai-26-10-2025>

42 Copyright and AI Reference Group <https://www.ag.gov.au/rights-and-protections/copyright/copyright-and-artificial-intelligence-reference-group-cairg>

Current licensing environment for AI

- 16.** The licensing market for AI training on restricted or privately held copyright works is new and growing. Below we discuss the current environment for direct licensing – where a right holder transacts directly with an AI developer or other party – and collective licensing – where rights from many different right holders are licensed or assigned to a CMO and licensed by them.

Direct licensing

- 17.** The private, commercial nature of contractual negotiations means that there is no single approach to how licences are negotiated. Parties negotiate and agree the terms between them. However, copyright licences will usually include common features such as the works covered, the purposes they can be used for, and the duration, cost, and territorial extent to which the licence applies.
- 18.** The current market for licensing copyright works to AI companies is somewhat opaque as contracts are a private matter. The types of work that feature in such deals will often be large catalogues of works owned or controlled by a single entity, such as a record label, news media organisation, or image library.
- 19.** One way to assess the licensing market is to track publicly announced deals. Analysis from the Centre for Regulation of the Creative Economy (CREATE)⁴³ showed that the news publishing industry entered into the most agreements with AI companies (68% of deals), followed by images (14%) and academic publishing (7%) between March 2023 and February 2025.
- 20.** The most frequently announced licensing deals came from dedicated AI companies (OpenAI and Perplexity), rather than traditional ‘Big Tech’ companies (Amazon, Apple, Google, Meta), potentially as the latter access data through other means, such as data mining their own platforms. Since the introduction of user prompts and the proliferation of daily users for chatbots, AI companies collect further data from their own services. Whilst these deals involve AI training licences, many are for access to restricted or privately held copyright works, rather than granting permission to use publicly accessible content.
- 21.** The same research notes 14 deals in 2024 where the content provider parent company is UK-based.⁴⁴ These were all for either news or academic publishing. Smaller news publishing deals for AI licensing have been reported at around £4 million per year,⁴⁵ though there have been significantly larger deals, such

⁴³ Centre for Regulation of the Creative Economy (2025) <https://www.create.ac.uk/blog/2025/02/24/the-ai-licensing-economy/>

⁴⁴ For the purpose of this analysis, we have included the deals connected with the Financial Times who have their head office in the UK but are owned by Japanese-based Nikkei.

⁴⁵ This includes general reporting that smaller news licensing deals are worth \$1-5m (The Verge <https://www.theverge.com/2024/1/4/24025409/openai-training-data-lowball-nyt-ai-copyright> OpenAI’s news publisher deals reportedly top out at \$5 million a year) as well as two datapoints in the Centre for Regulation of the Creative Economy dataset which both detailed \$10m deals over 2 years.

as that struck between OpenAI and News Corp which is estimated at around £40 million per year.⁴⁶ Some significant academic publishing licensing deals include one deal between UK-based Taylor & Francis worth at least \$10 million⁴⁷ and multiple deals with USA publisher Wiley reportedly worth \$21-\$23 million⁴⁸.

22. There have also been more recent announcements of AI licensing deals in other industries. For example, in the music industry, ElevenLabs⁴⁹ partnered with MERLIN, and Udio secured agreements with MERLIN,⁵⁰ Universal Music Group and Warner Music Group.⁵¹ The latter were agreed following legal disputes between the same companies. A licensing deal between The Walt Disney Company and OpenAI has also recently been announced.⁵²
23. The Publishers Association reported that UK academic publishers received £335 million in commercial TDM licensing revenue in 2021⁵³; and more recently described in their 2025 consultation response that they expected this to have grown substantially since then. This initial figure appears to be separate to the announced deals with AI companies (which were largely agreed from 2023 onwards) and suggests that there was a substantial TDM licensing market before these AI training agreements were in place. This revenue could reflect a collection of smaller purchases of TDM permissions by commercial research organisations.
24. There would appear to be a more limited prospect of a direct licensing market developing for individual right holders and SME right holders. In consultation responses, several groups representing primary creators said that they were unable to license their works to AI companies, and licensing examples tended to be from larger right holders. Reasons may include the size of catalogue that large right holders are able to control, access and other restrictions that they are able to place on the catalogue, and the ability to pursue legal action when needed.
25. Another factor will be the extent to which individual works or collections are substitutable with other content. Deals often relate to specialist datasets that cannot be easily substituted.
26. Some primary creators, including individual and SME right holders, will be beneficiaries of licensing deals by larger right holders. For example, Harper Collins is

46 The Wall Street Journal <https://www.wsj.com/business/media/openai-news-corp-strike-deal-23f186ba>

47 Informa <https://www.informa.com/globalassets/documents/investor-relations/2024/informa-plc---market-update.pdf>

48 Publishers weekly <https://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/95870-wiley-wraps-up-divestiture-program-looks-at-ai-opportunities.html>

49 Music Business Worldwide <https://www.musicbusinessworldwide.com/eleven-music-new-ai-rival-to-suno-launches-with-merlin-kobalt-licensing-deals-in-the-bag/>

50 Complete Music Update <https://completemusicupdate.com/ai-music-creation-platform-udio-signs-deal-with-indie-label-licensing-body-merlin/>

51 Music Business Worldwide <https://www.musicbusinessworldwide.com/udio-strikes-ai-licensing-deal-with-merlin-after-umg-and-warner-music-settlements/>

52 OpenAI <https://openai.com/index/disney-sora-agreement/>

53 Publishers Association <https://www.publishers.org.uk/wp-content/uploads/2022/08/22-8-Briefing-note-for-IPO-on-TDM.pdf>

reported to have sought consent from authors to provide licences for AI training in return for a \$5k payment per title, to be split equally between the author and publisher.⁵⁴

27. There is evidence that AI companies are willing to enter into agreements for publicly accessible but technologically restricted content, if the data is of high enough value to them and equivalent data cannot be obtained elsewhere. For example, Reddit changed its policy from allowing web crawlers to scrape content to requiring authorisation⁵⁵ and has struck major licensing deals with OpenAI⁵⁶ and Google⁵⁷; and they continue to take action against AI developers and web scrapers that have not agreed licensing deals⁵⁸.
28. These examples show the influence of website owners' adoption of technical access tools, and adherence to web crawler rules by developers and aggregators on licensing deals.
29. The evidence indicates that a market for direct licensing of restricted or privately held works to AI companies is developing, but several uncertain factors impact on the extent to which the licensing market can fully develop. Key themes highlighted by stakeholders include:
 - The legal position in the UK and other countries, notably the USA, and the extent to which content can be obtained without a licence.
 - The degree of transparency over which copyright works have been used in the AI development process.
 - AI companies cite transaction costs, including concerns about the difficulty of gaining the volume of permissions, particularly for SME firms and new entrants.

Collective licensing

30. Collective licensing is typically used where it would be impossible or impractical for direct licensing to take place due to the sheer number of both works and right holders involved. Individual and small right holders benefit from a stronger bargaining position, and licensees benefit from authorisation to use a wide repertoire of work and the efficiency of a simplified licensing transaction.
31. The collective rights management landscape in the UK is well-developed and covers many, but not all, sectors of the creative industries, including music, visual art, and literature. This reflects the commercial differences between sub-sectors of the

54 Publishers Weekly <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/96533-agents-authors-question-harpercollins-ai-deal.html>

55 The Keyword <https://www.thekeyword.co/news/reddit-ceo-calls-on-microsoft-to-pay-for-data-scraping#:~:text=Previously%2C%20Reddit%20allowed%20search%20engines,without%20proper%20attribution%20or%20compensation.>

56 Open AI <https://openai.com/index/openai-and-reddit-partnership/>

57 Reuters <https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22/>

58 AP <https://apnews.com/article/reddit-perplexity-ai-copyright-scraping-lawsuit-3ad8968550dd7e11bcd285a74fb6e2ff>

creative industries. For example, licensing of software is typically done directly and there hasn't been an identified need for collective licensing.

32. Right holders join a CMO voluntarily and instruct it to license their rights and collect their royalties. This means not all right holders of a particular sector are represented, even where a CMO exists. In 2024-25, CMOs in the UK represented over 600,000 right holders and distributed over £2.5 billion of royalties to them.
33. Collective licensing in the UK is market-led, and licences are developed to meet demand. It is yet to be known whether collective licensing for AI training will establish itself in the marketplace. This will depend on many factors including whether:
 - right holders mandate a CMO to negotiate licences on their behalf.⁵⁹
 - AI developers demonstrate interest in entering into collective licences.^{60 61 62}
 - all parties can agree on the terms of any licence negotiated.
34. Apart from the above-mentioned partnership activity by MERLIN, which is a CMO, there is no collective licence available in the UK for the training of generative-AI models. However, the UK's Copyright Licensing Agency, along with two of its member organisations (Publishers' Licensing Services and the Authors' Licensing and Collecting Society), has announced that it is currently developing a collective licence for generative-AI training.⁶³
35. Extended collective licensing (ECL) permits a licensing body such as a CMO to license on behalf of both member and non-member right holders (with the ability for right holders to opt-out should they choose). A CMO must demonstrate that it is significantly representative of right holders affected by an ECL scheme and must apply to the Secretary of State for approval to operate such a scheme, which will be subject to certain criteria and safeguards being met.
36. ECL streamlines permissions for the use of multitudes of works at scale. Therefore, if collective licensing for AI training is established, ECL may be a useful element, albeit there are no authorised ECL schemes in operation in the UK at this time.

59 Note that some right holders prefer exclusive licensing, for example the BPI, on behalf of the UK recorded music industry, noted within its consultation response that for sound recordings direct licensing "is emphatically the best way to create value"

60 DACS noted within its consultation response that it has faced "a variety of challenges to entering into licensing agreements with AI developers", citing that AI developers have not been willing to engage in negotiations and challenges due to a lack of transparency over what works are being used to train models.

61 Within its consultation response PRS for Music referenced that GEMA, a German performing rights society, offered a licensing proposal which was not taken up by AI developers leading to infringement proceedings.

62 Within its consultation response OpenAI noted that it did "not think that collective licensing should be encouraged nor is it appropriate"

63 Copyright Licensing Agency <https://cla.co.uk/development-of-cla-generative-ai-licence/>

Alternative licensing approaches

37. There are also alternative models to support the aggregation and licensing of content and data for AI training purposes.
38. As set out in Section F, technical tools and standards are being developed to support licensing and remuneration at the point of access to an online work. We are likely to see other technological licensing innovations as the AI market develops.
39. While not an alternative licensing approach, the government is supporting access to valuable datasets. On 26 January 2026, the government announced a pilot for establishing a Creative Content Exchange (CCE) as a trusted marketplace for selling, buying, licensing, and enabling permitted access to digitised cultural and creative assets and investment. This is part of the government's R&D Missions Accelerator Programme and includes 12 leading cultural institutions.

Views from the consultation

40. The consultation asked whether current AI licensing practice meets the needs of creators and performers, and the revenue and costs of such licensing. It also asked whether measures should be introduced to support good licensing practice, whether the government should have a role in encouraging specific types of licensing, and whether certain groups have specific licensing needs (questions 12 to 16).

Role of government encouraging specific types of licensing (including collective licensing)

41. The strong overall message is that the government should not seek to legislatively intervene in licensing, which should continue to be a commercial negotiation between the parties involved. This view came from both the creative industries and the AI sector. The same message was also strongly reiterated during the meeting of the technical working group on licensing.
42. For example, in its consultation response, UK Music said that “voluntary, market-driven licensing is a proven model that allows for rightsholders to negotiate their own terms and retain control of their works”, and government intervention in licensing was unnecessary. The British Copyright Council said that “We don’t see that the Govt should have any direct role in licensing beyond encouraging AI developers to license, individually or collectively”. TechUK noted its support of “an industry-led approach to licensing, ensuring commercial agreements develop in a way that reflects the diverse needs of both rights holders and AI developers”.
43. Creative sector stakeholders, including CMOs, noted that government should instead focus on ensuring the market conditions to enable licensing to flourish. Overwhelmingly, they argued for transparency requirements on AI developers, in the expectation that greater transparency would enable right holders to better license and enforce their rights. Several thought that an extra-territorial dimension to UK copyright law – such as the market access conditions in Regulation (EU) 2024/1689

(“EU AI Act”)⁶⁴ - would have a positive impact on encouraging licensing within the UK.

44. Several respondents noted that there were a growing number of direct licensing deals. A number of high-profile licensing deals have also been reported since the close of the consultation.
45. Technology sector representatives noted some challenges in licence negotiations, including ascertaining the right holder and determining a fair monetary value. They noted that much of the content available online did not contain rights management information to determine who the right holder was and how to contact them. Additionally, several noted that an individual work that had been used to train an AI model was likely to have a small or insignificant monetary transactional value, so the value to individual creators was likely to be small.
46. Many respondents from the AI and other technology sectors were against licensing for all uses of works (under the status quo, or stronger copyright) and favoured options for exceptions, such as those set out in Section B. Several made the point that an insistence on licensing of works that were accessible online was likely to chill AI innovation and may be unaffordable for new market entrants, SMEs and individuals.
47. Some technology sector representatives were positive that as the technology developed, the licensing opportunities would grow. There was also recognition that the licensing market was early stage, but developing, and that the ways in which data would be accessed were likely to become more innovative.

Licensing good practice

48. Some creator groups expressed concern that the benefits of licensing may not always flow to individual or SME right holders and primary creators.
49. For example, the Council of Music Makers said the government should ensure that consent is required from primary creators before their works can be used with AI, and that such consent should be “explicit, specific and meaningful”, not assumed through “generic terms in old agreements”. They proposed introduction of a new “exclusive right to synthesise” a work or performance with AI.
50. These groups also shared the views of the creative industries more widely that the primary measure to encourage the licensing of copyright works with AI was to ensure the application of copyright law through measures such as increased transparency about training materials.
51. A few technology sector representatives also observed that licensing was more likely to benefit larger right holders, rather than individual creators. Concerns were raised by groups including the Open Source Alliance and Startup Coalition, that licence-based access to creative content would naturally favour large companies and could lead to “walled gardens” of knowledge, reducing innovation by SME and individual developers.

64 The European Union AI Act: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

52. Users of AI tools, including researchers and libraries, also raised concerns that licensing could make it more difficult for users of AI tools to access knowledge, including researchers, and SME and individual users with fewer resources. They noted that some individual right holders, such as scientific academic authors may not share the same views as commercial authors and publishers, preferring open licensing of their works.
53. Several respondents mentioned mechanisms that could potentially stimulate the licensing market. These included a levy (e.g. Centre for IP Policy and Management, Bournemouth University), equitable remuneration (e.g. Musicians Union, the Centre for Regulation of the Creative Economy, British Equity Collecting Society), and R&D tax credit incentives to AI developers demonstrating compliance with copyright law or engaging in licensing agreements (e.g. Alliance for IP, Creative UK, Animation UK).

Collective licensing

54. There was acknowledgement within the creative sector that voluntary collective licensing could complement direct licensing for AI development purposes. Many noted that collective licensing was appropriate in some circumstances but not others, and that right holders should be free to determine the best solutions.
55. For example, the BPI, on behalf of the UK recorded music industry, noted that for sound recordings direct licensing “is emphatically the best way to create value” with collective licensing achieving only a fraction of the rates.
56. Several CMOs noted that the government should support licensing in all forms, whether individual or collective, but did not support legislative intervention. For example, PRS for Music noted that collective licensing does not “require a legislative role from government”, and PPL noted that “voluntary, market driven licensing is a proven model”.
57. Some AI developers also expressed opposition to legislative intervention, and others were sceptical about collective licensing more generally. Anthropic said “while compulsory or other collective licensing solutions could reduce transaction costs for licensing, they still raise many questions and potential problems from both policy and practical perspectives”.
58. Meta raised concern that “pressure to engage in collective licensing would go well beyond obligations in other jurisdictions [...] making the UK highly unattractive for AI developers”.
59. Anthropic noted that it was “important that policymakers not rush to pre-judge the end state of a fast-moving tech category and allow for those partnership conversations to proceed”.

Education and other uses

60. The consultation also asked about uses that affect specific sectors, bodies, and individuals. It asked specifically about copyright and AI in education, to help the Department of Education (DfE) consider the implications of AI for the education sector. The majority of respondents did not respond to this question as it was not relevant to them, but those who did respond prioritised respect for IP rights, and safety.

61. DfE has also taken forward its own process of engagement. It has been working closely with pupils, parents, and teachers to explore the appropriate framework for the use of pupil work, and any IP within it, as training material for the development of high-quality education AI tools.

Conclusions and proposals

62. The evolving licensing market demonstrates that creative materials can add value in the AI supply chain. The government believes right holders should be fairly remunerated for this value.
63. We propose not to intervene in the licensing market at this stage, as we do not consider there to be sufficient evidence to justify government intervention. Our proposals, mentioned earlier in this report, to work with industry experts to develop best practice on input transparency and to identify best practice on technical tools and standards may have positive outcomes in relation to licensing which will be kept under review.
64. We propose to keep market-led licensing approaches under review as the market for AI develops. This includes continuing to consider the likely effects on copyright owners, AI developers and users of AI systems who are individuals or SMEs, and whether the market is able to address issues affecting these groups, which were raised in the consultation. We will seek to learn as much as we can from past licensing deals and what they have achieved for the licensing ecosystem, while respecting and recognising the confidentiality requirements of these deals.
65. In relation to AI systems developed outside the UK, we propose to continue monitoring global developments and emerging international approaches like those proposed in India, and consider their potential effects on licensing in the UK. We also propose to continue monitoring litigation in the UK and elsewhere, including how secondary liability may apply to imported AI models placed on the UK market.
66. We propose to identify and assess further levers to support access to valuable datasets, including options to support the creation of, and access to, new public sector datasets – for example through the Creative Content Exchange (CCE). The CCE will test a range of commercial models for licensing, with the aim of launching an operational pilot platform by Summer 2026.
67. AI has the power to transform education by helping teachers focus on what they do best: teaching. DfE will continue to take forward its work to ensure the copyright framework allows the benefit of AI to be realised in education settings, protects the rights of pupils as creators of IP, and returns funds received back to the education sector. DfE is now assessing a range of options, to ensure pupil IP is used fairly.

Section H: Enforcement

1. This section considers ways of enforcing requirements and restrictions relating to the use of copyright works to develop AI systems including the accessing of copyright works for that purpose (for example by web crawlers). Current enforcement mechanisms are detailed, including relevant activity by regulators in this context. This analysis includes consideration of enforcement mechanisms which may be relevant in the case of AI systems developed outside the UK.
2. IP rights cannot have meaningful effect unless they are enforceable. The UK is internationally recognised for its strong framework for enforcing intellectual property (IP) rights, but AI may pose new challenges for enforcement in practice.
3. Many stakeholders have stated publicly that transparency obligations on AI developers are a prerequisite for effective enforcement, with some arguing UK copyright law should be applied to models trained overseas.
4. We believe enforcement should be effective, accessible and proportionate. Our approach is to ensure that the UK continues to have a competitive enforcement framework. If any new enforcement measures are considered, they must be accessible to right holders of all sizes including SMEs, providing effective redress while remaining proportionate to ensure that the UK is able to access the latest technology and innovation.
5. This section draws on responses to the consultation. Relevant consultation questions include those relating to infringement and enforcement (questions 7, 10, 11, 38 & 39). This section also draws on additional insights from stakeholders that have engaged with government since the consultation. Enforcement of copyright and wider regulation is also considered in the accompanying impact assessment.

Overview of existing mechanisms

6. Copyright law grants several exclusive rights⁶⁵ to a copyright owner, which enable them to control how their work is used. If someone performs one or more restricted acts in the UK without the copyright owner's consent – whether in relation to all or a substantial part of the protected work – it is considered copyright infringement, unless the law provides an exception for that activity. A discussion of how these rights and exceptions may apply to AI development is set out in Section B of this report.
7. Reproduction of any copyright work (or a substantial part thereof) to develop AI models in the UK, or the performance of any other restricted act, will require a licence from the right holders, unless a specific exception applies. This section concerns the enforcement of copyright, and related measures.
8. The UK is internationally recognised for its strong framework for enforcing intellectual property (IP) rights (for example, it is ranked second globally on

65 Exclusive rights: <https://www.gov.uk/guidance/the-rights-granted-by-copyright>

enforcement in the U.S. Chamber of Commerce’s international IP index⁶⁶). The framework, which supports proportionate and effective enforcement, is underpinned by flexible, technology-neutral legislation. This approach seeks to balance the needs of interested parties by encouraging and protecting innovation in the creative industries, media sectors as well as the emerging AI and technology sectors.

9. Although generative AI and the methods used to acquire training material, such as web scraping, are relatively new technological developments, the UK’s technology-neutral legislation provides various remedies to allow enforcement of IP rights which are applicable in this context.
10. Copyright is a private right which is normally enforced by individuals through civil litigation. In certain circumstances, however, copyright infringement may amount to a criminal offence.
11. In the UK, civil copyright claims are heard in either the Intellectual Property Enterprise Court⁶⁷ (“IPEC”) or the High Court, depending on the specifics of the claim.
12. The IPEC is a specialist court with simplified procedures, financial limits, and cost rules. It provides an accessible mechanism for smaller right holders, including individuals and SMEs, to seek redress, without requiring the resource and support that can be needed for normal litigation. The Court includes a small claims track⁶⁸ for cases worth £10,000 or less, where decisions are made without parties needing legal representation or being at risk of anything but very limited costs. Within the UK’s courts, a right holder can seek several remedies. These include damages to recover the costs of infringement or an injunction to compel an infringer to take certain actions, such as preventing further use of the copyright work.
13. To bring a case, several elements must be evidenced to the court. These include establishing that the work in question is protected by copyright, that the rights are owned by the claimant and that copyright has been infringed by the defendant. As there is no requirement for copyright to be registered, establishing ownership will require proof. This may not be straightforward, especially in cases where there are overlapping rights or shared authorship. It may also be necessary to consider whether relevant exceptions apply to the activity in question.
14. In their consultation responses, the technical working groups, and in wider discussions, creative industry stakeholders have argued that transparency obligations on AI developers requiring them to disclose what material has been used for training would significantly reduce the initial cost of pursuing enforcement actions. Getty Images in a press release following the High Court ruling in their case against Stability AI, said that “even well-resourced companies such as Getty Images

66 U.S. Chamber of Commerce’s international IP index <https://www.uschamber.com/intellectual-property/2025-ip-index>

67 Intellectual Property Enterprise Court: <https://www.judiciary.uk/courts-and-tribunals/business-and-property-courts/business-list-general-chancery/intellectual-property-list/intellectual-property-enterprise-court-ipecc/guide/>

68 Guide to the IPEC small claims track <https://www.gov.uk/government/publications/intellectual-property-enterprise-court-a-guide-to-small-claims/guide-to-the-intellectual-property-enterprise-court-small-claims-track>

face significant challenges in protecting their creative works given the lack of transparency requirements. We invested millions of pounds to reach this point with only one provider that we need to continue to pursue in another venue. We urge governments, including the UK, to establish stronger transparency rules which are essential to prevent costly legal battles and to allow creators to protect their rights.”

15. With respect to AI training conducted abroad, the UK is a signatory to multiple international agreements⁶⁹ that ensure copyright material produced by UK nationals or residents, which falls within the scope of these conventions, is automatically protected under the national laws of each member country. In the case of potential infringement abroad the copyright owner will need to consider the local laws to determine if actions taken to train AI models have infringed their copyright in that jurisdiction.
16. Under UK law copyright in a work can be infringed “by a person who, without the licence of the copyright owner, imports into the United Kingdom, otherwise than for his private and domestic use, an article which is, and which he knows or has reason to believe is, an infringing copy of the work”⁷⁰.
17. As noted earlier in this report, the recent judgment in Getty Images vs Stability AI found that an AI model, although intangible, may comprise an ‘article’ and as such, making a model trained overseas available in the UK market may infringe copyright if that model includes an infringing copy. In the Getty case the Judge found that the model in question did not retain any copies of works, and as such was not an infringing copy (and further that no importation occurred unless the model was available for download by UK users as opposed to being served remotely), but this finding was fact specific and is the subject of an appeal to the Court of Appeal (likely to be heard in 2026). To reinforce the fact specific nature of this ruling, a recent German court decision in GEMA v OpenAI found that an LLM in that case did retain copies of song lyrics it had been trained on.
18. Where an AI model is used to generate material that reproduces all or a substantial part of a copyright work without permission this may also comprise an infringement of copyright if there is no relevant exception and no licence is in place. This act of infringement occurs at the output stage and may also create infringement through any subsequent dealing. Any action would be taken against the persons responsible for these respective acts, and depending on the circumstances, the user, the provider of the AI system and any person dealing with infringing content after it has been created may all be liable. Enforcement action against such infringement is available as for other infringements of copyright and would usually be pursued by the right holder through the civil courts.
19. AI developers and service providers often take steps, both during training and at inference, to avoid outputting protected works. This may include de-duplication, to help avoid “overfitting” during training (see Section B), as well as filtering certain prompts and outputs.

69 Intellectual Property Office guidance on protecting your copyright abroad <https://www.gov.uk/guidance/protecting-your-copyright-abroad>

70 Section 22 Copyright, Designs and Patents Act 1988 <https://www.legislation.gov.uk/ukpga/1988/48/section/22>

20. The UK also has established notice and takedown systems, where an online host may be requested to take down copies of infringing material or become liable for that infringement if they do not do so. These processes would be applicable in the case of copyright infringing AI-generated content that is posted or shared on online platforms.
21. Where infringement takes place for commercial benefit, or to such an extent as to affect prejudicially the owner of the copyright, this may comprise a criminal offence. In these circumstances there are criminal enforcement options through private criminal prosecution or via law enforcement agencies. The most active law enforcement body focused on this activity is the Police Intellectual Property Crime Unit (PIPCU) operated by City of London Police. PIPCU works closely with right holders to investigate and enforce against online infringements. In addition, UK Trading Standards, UK police forces and Border Force work together to disrupt large-scale criminal IP infringement.

The role of a regulator

22. Copyright is a private right, which means it is not directly enforced by a regulator. There is no regulator in the UK which has a role in the enforcement of copyright. The IPO is responsible for copyright policy but has only a limited regulatory role in relation to collective licensing. The Copyright Tribunal⁷¹ has a role adjudicating collective licensing disputes but does not have any role in considering issues of infringement.
23. Previous sections of this report have noted the benefits of AI developers adopting good practice in areas including transparency and technical standards. Good practice could also cover copyright compliance, such as the avoidance of infringing outputs. Regulatory oversight would be one way to ensure such practice is adopted. Under such a model a body could be assigned to oversee the adoption of, and compliance with, a code of practice, with the ability to impose sanctions for non-compliance.
24. The EU has taken such a regulatory approach in its AI Act and Code of Practice. During the passage of the Data (Use and Access) Act 2025 (the D(UA) Act), many Parliamentarians advocated a similar regulatory approach in the UK. However, no specific regulator with this remit currently exists.
25. The Information Commissioner's Office (ICO) has the power to regulate some data access activities. It was proposed in a non-government amendment during the passage of the D(UA) Act that the ICO should act as a regulator in relation to AI and copyright. However, it does not currently have a role in issues relating to copyright and AI, or funding allocated for this purpose.
26. Ofcom's primary function is to regulate communications in the interests of consumers and citizens. This includes the oversight of some online activity, including compliance with the Online Safety Act 2023⁷². However, its remit does not extend to

71 Copyright Tribunal website <https://www.gov.uk/government/organisations/copyright-tribunal>

72 Online Safety Act 2023 <https://www.legislation.gov.uk/ukpga/2023/50>

copyright, or the matters covered by this report. Any new powers would need to be prescribed in legislation and funded.

27. The Competitions and Markets Authority (CMA) has a role under the Digital Markets, Competition and Consumers Act 2024⁷³ to regulate firms operating in digital markets who have strategic market status. The powers available to the CMA include the imposition of conduct requirements on designated undertakings (firms). These powers are intended to address specific issues and risks arising from a firm's strategic market status, and do not extend to general regulation of a sector and all the players operating within it.
28. Creation of a new regulator or placing additional duties onto an existing regulator would however incur significant costs and would require funding. Providers of general-purpose AI models to the EU market already have to comply with regulation on transparency and other matters. Additional regulation in the UK would add further complexity to the regulatory landscape, and the impact of any such measures would need to be carefully assessed.

Market access and extra-territoriality

29. During the passage of the D(UA) Act, non-government amendments were laid which aimed to extend the scope of UK copyright law and related transparency measures to models trained overseas. Although those amendments did not pass into law, there was considerable debate about the issue of extra-territorial effect and the extent to which it would be possible or desirable for the UK to ensure that models developed elsewhere are only able to enter the UK market if they are developed in line with UK copyright or other regulatory rules.
30. A market access approach such as this would be similar to the regulatory approach used in other sectors of the economy (pharmaceuticals, electronics, automotive etc). The EU has taken this approach in Regulation (EU) 2024/1689 ("EU AI Act"), applying transparency and other rules to certain types of AI model and service on a market-access basis. Such an approach could be taken in the UK, which would mean that an AI model or service (or certain types of model and service) could not be marketed in the UK unless compliant with UK rules.
31. Market access restrictions could be complex to enforce and would be likely to require a regulator to underpin their effectiveness. The EU's approach is still in its infancy and its efficacy in practice is still unknown.
32. A market-access approach would also need to be carefully balanced to ensure that the UK remains a competitive jurisdiction for the development and deployment of AI. Earlier in this report we discussed potential effects on AI systems developed outside the UK if UK copyright law were extended to them. Similar considerations would need to be taken into account if transparency and other rules were applied on a market access basis. Rules which place disproportionate burdens on AI developers could affect the availability of AI systems on the UK market. Care would need to be taken to get the balance right.

⁷³ Digital Markets, Competition and Consumers Act 2024 <https://www.legislation.gov.uk/ukpga/2024/13/contents>

Views from the consultation

33. While the government considers the UK's framework for enforcing copyright to be generally well placed to adapt to new technological innovations, it invited views on whether action should be taken to improve right holders' control over their works, which was one of the objectives of the consultation. We welcome any opportunity to ensure our enforcement framework remains world class.
34. The consultation sought views on how to enhance control and legal compliance in the context of options 1 (stronger copyright) and 3 (the opt-out exception). The consultation cited increased transparency as a way to ensure copyright is complied with and can be enforced, and asked questions about this. It also sought views on compliance with technical standards to enhance right holder control (questions 7, 10 and 11). It sought views on the current approach to liability in AI-generated outputs, enforcement in relation to outputs, and what steps AI providers should take to avoid copyright infringing outputs (questions 38 and 39).
35. Several of these aspects have been covered by other sections in this report. This section specifically considers enforcement aspects. Below, we summarise general views provided to the consultation on enforcement, as well as views on specific enforcement-related questions.

General comments on enforcement

36. In general, right holders supported measures to make it easier to enforce their copyright when works are used by AI developers without permission. The primary issue cited by many was the lack of transparency over training material.
37. For example, the Alliance for IP, responding on AI-generated outputs (question 38), said the law was clear in relation to such outputs but that enforcement "is clearly challenging" given the lack of transparency. They said the solution was for AI developers to seek consent and licences, but that legal recourse "clearly is not available for many smaller right holders".
38. The British Copyright Council, echoing the views of many creator groups, noted the importance of the UK's already strong copyright and enforcement framework, and urged the government to ensure that any penalties should be "set at a sufficiently high level to deter infringement".
39. In the context of the opt-out exception (option 3), and compliance with technical standards, many respondents favoured strong enforcement including additional penalties. The same respondents tended to support maintaining or strengthening copyright law, and licensing of copyright works.
40. Responding to the question on non-compliance with opt-outs (question 7) and technical standards (question 10) the majority favoured financial remedies, including damages, and many noted that the usual remedies under copyright law would apply. A large proportion also felt that there should be operational sanctions for non-compliance, such as a duty to remove works from training sets and a trained model. Others indicated support for criminal prosecution and penalties for non-compliance.

41. The Alliance for IP suggested that systems for rights reservation “must have the same standing in law as Technical Protection Measures” and that “subverting such measures must therefore have criminal penalties.”
42. Some respondents called for UK copyright rules to be extended to cover actions taking place outside the UK. For example, the News Media Association (NMA) said that government should “explicitly expand it [the copyright framework] to cover all GAI models [generative AI models] linked to the UK”. The NMA also called for a suitable regulator to be empowered to levy fines for non-compliance with transparency obligations.
43. Many respondents advocated a government role in ensuring compliance with transparency obligations (question 22) and in technical standards adoption (question 11), with others indicating they believed the government should have a role in ensuring compliance more broadly. The Professional Publishers Association stated that “a regulator should have oversight and enforcement powers, including the ability to impose fines for non-compliance”.
44. On AI-generated outputs, the consultation asked whether the current approach to liability allows effective enforcement of copyright. 41% of respondents in the online survey answered “no” to this question, (3% answered “yes” and the rest did not answer). A common sentiment (which largely reflects the most common campaign response) was support for both AI developers and users being liable for copyright-infringing outputs, but a preference for this falling on AI developers. Of the respondents who answered “yes”, some indicated that a lack of transparency still limits enforcement and greater transparency would aid enforcement in relation to outputs.
45. On the steps that AI providers should take to avoid copyright infringing outputs (question 39), the most common theme was the avoidance of infringing inputs, for example training data, discussed earlier in the consultation. Beyond this, the most common themes identified related to scanning or checking AI outputs; scanning or checking AI prompts; and measures that can be put in place to educate or inform users of how their prompts and resultant outputs can be used in relation to copyright law. Another theme was the need for cooperation with right holders to find solutions. There were again some calls for greater transparency and enforcement.
46. Some respondents to the consultation sought greater clarity from the government on what constitutes an infringement in the context of AI-generated content and who should be liable. Some suggested by issuing guidelines on the threshold for infringement and where liability rests, the Government could improve how AI tools are used and help reduce the volume of AI-generated infringing content.

Conclusions and proposals

47. Throughout the consultation, and during our subsequent engagement with stakeholders, the government has been clear about the importance of effective, proportionate copyright enforcement. We propose to continue working with partners across relevant sectors, as well as with law enforcement and the judiciary, to help ensure the UK enforcement framework provides effective and accessible routes to redress for those who need it in relation to infringement in the context of AI. This work will consider enforcement issues for AI systems developed in and outside the

UK, including the likely effect on copyright owners, developers and users who are individuals, micro businesses, small businesses or medium-sized businesses.

48. The range of views provided by consultation respondents and by stakeholders during technical working groups has provided valuable insights into our ongoing evaluation of how our enforcement framework meets our stated goals in the context of this technological development.
49. The government considers the UK's framework for enforcing copyright to be effective and capable of adapting to developments in AI. However, while wider policy on AI and copyright continues to be developed, we propose a programme of further work to consider ways of enforcing requirements and restrictions relating to the use of copyright works to develop AI systems and the accessing of copyright works for that purpose, both for AI systems developed in and outside the UK. This work will include considering the likely effect of any proposals that are made on copyright owners, developers and users who are individuals, SMEs and micro businesses (including individual creators) with fewer resources.
50. As part of this work, we propose to consider what further actions the government, industry, judiciary, and law enforcement can take to help mitigate these barriers.
51. As noted previously in this report, we propose that we do not introduce regulatory oversight on transparency and other measures at this time. Accordingly, we propose that no new regulator should be created specifically to oversee matters of AI as it relates to copyright, and that no regulatory duties relating to these matters be imposed on existing regulators.
52. We propose to continue monitoring and assessing regulatory and enforcement approaches taken in other jurisdictions, including the EU, paying specific attention to building a clear evidence base regarding the effect of different enforcement mechanisms on the development and deployment of AI models, and the accessibility of effective redress for right holders of all sizes.

Section I: Computer-generated works

1. In the UK, copyright protection is available for any literary, dramatic, musical, or artistic work which is “generated by a computer in circumstances such that there is no human author”.⁷⁴ This computer-generated works (CGWs) protection lasts for 50 years from the date of creation. In terms of copyright ownership for CGWs, the “author” of such a work is deemed to be the person “by whom the arrangements necessary for the creation of the work are undertaken”. In the case of a general-purpose AI which generates output in response to a user prompt, the “author” will usually be the person who inputted the prompt.
2. This section considers the existing copyright protection that is given to CGWs, including the types of protection available in UK law, international comparisons, stakeholder views, and proposed next steps. The consultation included several questions relating to CGWs (questions 30-37). This section draws on responses to these questions as well as other sources and focuses on the extent to which the protection continues to be justified.

Rationale for computer-generated works protection

3. The CGW provisions were introduced in the Copyright, Designs and Patents Act 1988 (CDPA), as enacted. During the Parliamentary passage of the Copyright, Designs and Patents Bill, Lord Young of Graffham, the Secretary of State for Trade and Industry, described it as “the first copyright legislation anywhere in the world which attempts to deal specifically with the advent of artificial intelligence”. The aim was to “allow investment in artificial intelligence systems, in the future, to be made with confidence”.⁷⁵
4. As well as content generated by AI, the provisions were intended to cover content such as statistical data, weather maps, and outputs from expert systems. These were felt to be valuable but, at the time, did not clearly benefit from copyright protection.
5. However, it is not clear that this rationale for CGW protection still applies. While producing CGWs may have been challenging and costly in the 1980s, AI has since developed significantly to where CGWs can be produced in large quantities without the same challenge or cost.
6. Several of the other content types which were intended to be covered, such as compilations of weather data, are now likely to be protected by other rights, such as the database right, introduced into UK law via an EU Directive in the 1990s.⁷⁶

⁷⁴ Section 9(3) Copyright, Designs and Patents Act Copyright, Designs and Patents Act 1988

⁷⁵ Copyright, Designs and Patents Bill HI Volume 489: debated on Thursday 12 November 1987: <https://hansard.parliament.uk/Lords/1987-11-12/debates/9b959a7b-172a-4e28-8676-1a6747b0f370/CopyrightDesignsAndPatentsBillHI>

⁷⁶ Directive 96/9/EC: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:31996L0009>

7. A key rationale for copyright is to remedy market failure by incentivising the production of creative output. However, it is not clear that this market failure exists in the case of wholly AI-generated creative output, where the marginal cost of production is very low. As such, the consultation on copyright and AI sought views on the future of this right.

Other forms of copyright protection for AI outputs

8. In addition to CGW protection, copyright may apply to the outputs of AI systems in other ways, depending on the circumstances of their creation.
9. First, an output which is sufficiently original may qualify as a literary, dramatic, musical, or artistic work (known as “authorial works”). These are defined in sections 3 and 4 of the CDPA.
10. This is the same copyright that applies where a person has created a work using any other means. In order to be so protected, such a work must be original in the sense that it expresses the intellectual creation of its author, reflecting their personality and creative choices.⁷⁷
11. In general, any work made using AI as a tool, but where the creative expression in the work comes from a human creator, will meet this originality requirement. An example could be an original photograph which is taken by a human photographer and edited using an AI tool, or an original literary work which is edited with help from an AI assistant. We refer to such works in this section as “AI-assisted works”. They are protected in similar terms in the EU, USA, and many other countries.
12. A second type of protection applies to “entrepreneurial works”. This protection applies to sound recordings, films, broadcasts, and typographical arrangements (defined in sections 5A, 5B, 6, and 8 CDPA). Copyright accrues to the producers, broadcasters, or publishers of such works, and aims to protect and reward their investment in them. Such protection applies regardless of human creativity, which means, for example, a recording of music is protected in similar terms to a recording of birdsong. It will also apply to a recording of music, film, etc. generated by an AI system. Sound recordings, films, and broadcasts are protected to differing degrees in different countries.

Legal scope of computer-generated works protection

13. Section 9(3) CDPA sets out the provision for CGWs: “In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken”.
14. There appears to be a legal contradiction within section 9(3) which leads to uncertainty about its interpretation. This is because the provision applies only to

⁷⁷ THJ Systems Ltd & Anor v Sheridan & Anor [2023] EWCA Civ 1354 (20 November 2023) at [15] and [23]: <https://www.bailii.org/ew/cases/EWCA/Civ/2023/1354.html>

literary, dramatic, musical, and artistic works which are original. The legal test for originality, as it has evolved since the CDPA was made law, is that a work must be an “author’s own intellectual creation” which is the expression of their creative choices and reflects their “personal touch”. This test is very much associated with human qualities, suggesting that a work created by a non-human could not be “original”. However, section 9(3) only applies to works “without a human author”.

15. This contradiction has led some to question whether the provision could ever apply in practice. In our view, it is unlikely that a court would conclude that it can never apply, as Parliament clearly intended the provision to have an effect. But it is unclear in the absence of case law how an “original” yet wholly machine-authored work would be defined.

International comparisons

16. Most jurisdictions do not provide specific protection for CGWs. This includes the USA and China, which both experience higher investment in AI than the UK. This may support the notion that incentives to develop AI are not related to the existence of equivalent CGW protection for AI outputs. A small number of other countries, including India, New Zealand, Hong Kong and Singapore do provide protection similar to that in the UK. However, like the UK, the legislation in these countries has been largely untested in the courts.
17. Below we consider the situation in the EU, the USA, and China.

The European Union

18. The EU requires its member states to provide copyright in original works, and related rights in phonograms (sound recordings), film fixations and broadcasts. These correspond to the “authorial” and “entrepreneurial” copyright protection available in UK law, as described above.
19. To qualify for protection as an authorial work, a work must be original in the sense that it is the author’s own intellectual creation, expressing their free and creative choices. The UK inherited this originality standard from EU law. As such, copyright in authorial works, including AI-assisted works, is available in the EU on a similar basis as it is in the UK.
20. The EU does not provide for specific protection for computer-generated works without a human author. The originality standard developed by the Court of Justice of the EU is tied to human concepts such as creativity and personal expression. This is explored further in the European Parliament’s ‘Generative AI and Copyright’ study.⁷⁸ As such, it is unlikely that such wholly AI-generated outputs will benefit from copyright protection in the EU.

⁷⁸ European Parliament JURI Committee Study ‘Generative AI and Copyright: Training, Creation, Regulation, July 2025: [https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU\(2025\)774095_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU(2025)774095_EN.pdf)

The United States of America

21. Like other countries, copyright protection in the USA arises automatically upon creation of a work. However, protection is enhanced by registering a work with the U.S. Copyright Office (USCO). This supports enforcement of exclusive rights before the courts.
22. The USCO may decline to register a work if it does not qualify for protection. Because of this, it has had to grapple with the extent to which generative AI tools can be used to create registrable copyright works.
23. The USA follows a similar human authorship requirement to the EU. Under USA law, copyright is only granted to a natural person exercising creative choices. This means copyright protection will be refused if a human did not create a work. This includes when a machine operates autonomously or randomly, without meaningful human input or intervention.^{79,80}
24. As in UK and EU law, where a human has authored a work with AI assistance, that work will be capable of protection in the USA, if the originality standard is met. To register a work that contains more than a de minimis amount of AI-generated material, the application must disclose that information and describe the human author's contribution. The USCO has registered hundreds of works that incorporate some form of AI-generated material.
25. The USCO January 2025 report on the copyrightability of AI outputs explored these themes in more detail. It confirmed that, where AI-generated content is arranged in a sufficiently creative way that "the resulting work as a whole constitutes an original work of authorship", or AI-generated works are modified to such a degree that the modifications meet the originality standard, these works will be protected. It includes an example of an AI tool that allows a user to control the selection and placement of individual creative elements (for example, AI tools that permit musicians and sound engineers to modify recordings or tools that enable film editors to edit film). To conclude, the USA also does not provide for specific CGW protection for wholly AI-generated content without a human author.

China

26. China does not have specific provisions for CGWs. However, a series of cases seeking copyright protection for CGWs have come before the courts. In some of these cases the courts have ruled that the AI-generated outputs should be entitled to copyright protection.
27. In the case of *Li Yunkai v. Liu Yuanchun*, the plaintiff used a generative AI tool to create an image which they shared online. The image was then shared online by the defendant with the plaintiff claiming infringement. The court ruled that this was

79 U.S. Copyright Office, Compendium of U.S. Copyright Office Practices 101, 306, 312.2 (3d ed. 2021)

80 Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190 (Mar. 16, 2023).

infringement on the basis that the plaintiff had determined certain details of the image through the prompts used.⁸¹

28. In another case, the court ruled in favour of the plaintiff who had created an image using generative AI and then made edits using editing software. The image, and an adaptation of the image, was then used by two businesses without the permission of the plaintiff. The court ruled in favour of the plaintiff who received financial compensation from both defendants.⁸²

Views from the consultation

29. The consultation sought views on the UK's existing provisions for protecting CGWs. Our policy options for CGWs were assessed against three objectives:

- clarity over what is and is not protected by copyright;
- whether they incentivise and reward creative output without over-regulating; and
- whether they achieve a balance between encouraging human creativity without hindering technological development.

The consultation also sought to understand how right holders are using CGW protection, and what their economic impact has been.

30. Three potential options were presented:

- **Option 0: No legal change, maintain the current provisions**

This would be our approach if there was evidence that CGW protection is necessary to encourage the production of outputs by generative AI or similar tools, and that any legal ambiguity is likely to be of little impact.

- **Option 1: reform the current protection to clarify its scope**

Under this option, we would clarify the existing copyright protection for CGWs. Protection of AI outputs by “traditional” copyright (authorial and entrepreneurial) would be unaffected.

- **Option 2: remove specific protection for CGWs**

Under this option we would remove the specific protection provided to CGWs by section 9(3) CDPA. Works which are AI-assisted, but which nonetheless exhibit human creativity would continue to be protected. AI-generated music and video could continue to be protected as sound recordings, films, and broadcasts. However, there would be some impact on text and image-based content generated

81 Bird & Bird article ‘Copyright Protection for AI generated works’: <https://www.twobirds.com/en/insights/2024/china/copyright-protection-for-ai-generated-works-recent-developments>

82 DLA Piper article ‘Another Chinese court finds that AI-generated images can be protected by copyright...’: <https://www.dlapiper.com/en-at/insights/publications/2025/05/another-chinese-court-finds-that-ai-generated-images-can-be-protected-by-copyright>

purely by AI without a human creator.

31. The Government's preferred option was to remove CGWs protection unless the consultation responses provided sufficient evidence of its positive effects.
32. The consultation asked questions about several aspects of CGWs protection, including how it should be interpreted and whether it is currently being relied on. It also asked whether the provisions should be clarified, amended, or removed, and the potential impacts of doing this.

Overall sentiment

33. Less than half of the total consultation respondents answered the CGWs questions.
34. The majority of those who did respond were not in favour of maintaining the current CGWs protection (78% of online survey respondents) and agreed that there should not be protection for works created solely by AI. A common theme identified in the responses was that CGWs compete with human works. Removing copyright in CGWs would mean humans not having to compete with AI outputs. Respondents to this question also pointed to positive impacts that removing the CGW right would have on right holders, including financial impacts and control of their works. A very small number of respondents indicated here that there would be a negative impact on AI companies or users of AI. This was on the basis that AI systems could be less appealing if users could not protect the outputs they created. A significant proportion expressed no strong interest in the outcome either way.
35. Many individuals strongly supported removing the rights, expressing concern that the rise of generative AI coupled with CGWs protection risked flooding the market with AI-generated content and undermining human creators. A common theme, particularly from those in the creative sector, was a feeling that this protection risks devaluing human artistic expression in the age of AI.
36. Some responses stated that they simply did not use CGWs protection or see any substantial value in retaining the existing CDPA provisions. Right holders and the creative industries generally supported CGWs removal given that AI-assisted works would remain protected by copyright. Relatively few respondents expressed views relating to keeping the current protection in place.
37. Few AI sector responses provided detailed views on CGWs protection. However, some AI companies suggested they would be unaffected by CGWs removal. For example, Open AI said they would be unaffected as they "do not claim copyright over generated outputs".

Legal considerations and challenges

38. Out of those who responded to question 32 in the online survey, 87% indicated CGWs legislation would benefit from legal clarity. The themes of several of those responses were repeated from earlier questions such as general support for removing the CGWs right and support for creators having a share of the CGWs rights (where the respective models are trained on their creative works). A less common theme was the view that the originality requirement should be clarified, whilst another less common view expressed was that the law could be clarified by introducing transparency measures.

39. The consultation highlighted an apparent contradiction between section 9(3) CDPA, which applies to original works without a human author, and the originality standard, which is associated with human qualities such as personality. It asked how this issue might be addressed, should this type of protection be maintained.
40. A number of respondents gave general support to the need for some form of legal clarity in how the current legislation functions. For example, a trade body representing the creative industries said ‘If the government intends to retain [CGWs], then concerns over lack of clarity in the way that the authorship provision links to the authorship and ownership provisions applied to types of copyright works recognised under Berne and TRIPS must be considered’.

Potential reform

41. A number of respondents sought clarity on the current CGWs provisions with some sharing specific ideas for reform. Most of these reform ideas aimed to support the creative industries rather than the AI industry. The most popular reply to the questions on reform was that wholly AI-generated and autonomous outputs should not be covered by copyright, but AI-assisted creations should. However, as described above, this is essentially what would happen should CGWs protection be removed, as AI-assisted works would be eligible for copyright protection like any other original work.
42. Other responses set out more ambitious ideas for reform that would also favour the creative industries. For example, of the 3,465 survey respondents who gave open text views on question 31, a common theme was that creators should have a share of the rights and/or remuneration of CGWs, particularly where these outputs came from AI models that were trained on their creative inputs.
43. Similarly, some respondents suggested that AI users and developers should be required to disclose what works have been used in the creation of a CGW in order for an output to benefit from protection and as a form of transparency. For example, one academic research centre suggested “Require creators and AI developers to disclose when AI has been used in the creative process, ensuring transparency for consumers and rights holders”.
44. Respondents also proposed reforms to the scope of the current CGWs protection. One suggestion was to reform it to operate like an entrepreneurial right. This would protect the fixation of a CGW in a similar manner to sound recordings or broadcasts.
45. Other respondents suggested, potentially in combination with making CGWs operate as an entrepreneurial right, reducing the duration of CGWs protection. There was a range of suggestions for reduced duration, generally between 2 and 10 years. These respondents considered a limited term of protection would make the protection offered to CGWs more commensurate to the level of effort required to produce a generative AI output.

Economic considerations

46. Question 36 of the consultation asked what the economic impact of removing CGWs protection would be. A number of respondents used their responses to reiterate their position on the removal of the protection without providing any specific views on the economic impact. Some did provide views on the economic impacts, but in line with

the general lack of evidence about the use of CGWs protection, those responses were often not supported by clear evidence.

47. Where respondents indicated that removing CGWs protection would have a positive economic impact for the creative industries, this was based on the idea that the increased production of CGWs would decrease demand for human authored copyright works in the creative industries. Given the relative ease with which AI outputs can be generated, some respondents referred to market saturation, with AI-generated CGWs flooding the market in increasing numbers and reducing the demand for human created works.
48. Some respondents also highlighted a possible negative economic impact on the AI industry from removing CGWs protection. Some suggested removing CGWs would disincentivise AI development. Several of these responses framed disincentivising the AI industry as a positive, and consistent with their wider pro-creative industry stances.
49. Several respondents did provide quantitative evidence to support their views that CGWs protection should be removed. However, these were often based on speculative and estimated figures. For example, a respondent from the creative tech sector said 'if AI-generated content contributes 20%-30% of the output in creative industries like gaming or publishing, removing protections could disrupt those markets and lead to losses in licensing income'. That example may also be referring to AI assisted works which may qualify as authorial works. Some respondents provided similarly speculative figures to suggest that CGW protection will contribute to the growth of the AI industry in the UK.
50. Question 37 asked respondents to indicate how the removal of CGWs protection would affect their organisation, with options ranging from a significant positive effect to a significant negative effect. The response rate to this question was relatively small (34%) although of those, 41% indicated it would have a minor or significant negative effect. The most common theme of the written responses to question 37 was that removing the CGWs protection would cause confusion, and a better option could be to reform the protection.

Conclusions and proposals

51. In the consultation paper, the government indicated that, should insufficient evidence emerge of overall positive effects from CGWs protection, our preference would be to remove it.
52. The responses to the consultation show minimal evidence that CGWs protection is being used or has significant economic effect. The majority of respondents who engaged with the CGWs questions supported removal. While this partly reflects the high volume of creative industry responses, there was also limited support for maintaining CGWs protection from the AI sector.
53. We agree that copyright should incentivise and protect human creativity. There is minimal evidence that protection for CWGs is actively used, or that it has a material impact on creativity and innovation. We propose to continue to monitor the use and impact of this protection. However, in the absence of evidence of its ongoing value, we propose that it should be removed.

Section J: Digital replicas

1. This section discusses the use of AI to replicate or mimic the appearance or voice of individuals - variously referred to as “digital replicas” or “deepfakes” - and related practices.
2. The following analysis considers recent developments in the ease and realism with which replicas can be generated, the increasing impacts on those being depicted and the wider public, and how existing UK law applies.
3. In summary, digital replicas increasingly impact individuals and institutions across society, beyond just the creative industries. Realistic impersonation was once relatively rare, focused on criminal purposes or well-known personalities, and addressable through criminal offences or civil lawsuits. But developments in AI are making digital replicas more commonplace, and existing laws may not protect the public sufficiently as they increasingly experience imitation in everyday life without their consent.
4. The government therefore proposes to explore a range of options to help clarify when digital replicas are legitimate, and to help guard against unacceptable imitation. This will include exploring the case for greater commercial protections, and whether these should form part of any wider safeguards or rights to personality within the UK.
5. The government’s consultation on AI and Copyright included some initial questions on this issue. This was focused on the commercial and artistic considerations around digital replicas and did not consider in depth issues around criminal misuse. Relevant responses are also summarised in this section.

Overview of impacts from digital replicas

6. Digital replicas, like AI-generated media more broadly, have beneficial uses that can improve lives, support livelihoods, and boost wellbeing. Examples include naturalistic dubbing or translation that preserve a performer’s distinctive voice; reduced need for heavy cosmetics, risky stunt sequences or extensive travel; assistive technologies such as personalised synthetic voices for disabled people; more realistic training and simulation in healthcare and other settings; richer peer-to-peer digital communication; and productivity gains from ‘digital doubles’ in meetings and routine appearances.
7. At their best, this technology can empower people to participate more fully in the full breadth of human experience and make creative and professional opportunities more accessible.
8. However, the imitation of individuals without their consent or for illegal purposes is also a growing problem. Non-consensual replicas can impact both the person depicted and the wider public, in ways ranging from reputational harm and deception to criminal activity and violations of personal dignity. We have, in recent months, seen the use of commercially available tools to generate non-consensual replicas at scale that constitute both child sexual abuse and intimate image abuse.

9. Digital replicas can negatively impact people and businesses in a number of ways, including:
 - **Commercial imitation.** Replicas can be used to mimic the distinctive voice or likeness of an individual to imitate their work without permission, undermining contractual agreements, compensation and livelihoods. This was the focus of the digital replicas copyright and AI consultation section.
 - **Reputational harm.** Replicas can falsely depict individuals performing or participating in events and activities that cause them reputational harm, leading to impacts on relationships, self-esteem and financial earnings. This can extend to harassment, blackmail and being falsely accused of a crime.
 - **Deception and manipulation.** Replicas can deceive or manipulate audiences and institutions (such as the criminal justice system) at scale. They can alter beliefs, erode trust, radicalise individuals and cause people to act or make choices based on false information. The misuses of this technology include defrauding people and businesses, attempting to influence how people vote, and spreading mis- and disinformation.
 - **Technology facilitated sexual abuse.** Replicas can harm those imitated by depicting them in ways, settings or contexts that violate their consent and personal dignity – such as, but not limited to, Intimate Image Abuse (IIA), and Child Sexual Abuse Material (CSAM). These harms can have a devastating and lifelong impact on those depicted. The creation and dissemination of such material is a criminal offence and can also be used to facilitate other criminal acts, such as sexual extortion. These violations can also impact friends and family, including where depictions are created of deceased relatives.
 - **Production of criminal material.** Content including digital replicas can itself constitute criminal material, for example if it depicts CSAM or the endorsements of proscribed terrorist organisations. AI-generated criminal material creates an additional burden on law enforcement when attempting to identify whether a person or child depicted in a replica is a real victim or synthetic, which risks real victims not being identified and protected.
10. The consultation did not explore the harms from digital replicas, such as illegal harms, or those that may not meet the legal threshold for illegality but, nonetheless, cause real harm.

Application of existing UK law

11. UK law does not provide a general personal image right. Instead, it offers a patchwork of protections that can apply in certain circumstances – including data protection, online safety regulation, criminal offences, civil claims (such as defamation), and IP-related rights, such as trade marks and passing off, for people in the public realm. Historically, this patchwork reflected the uncommon nature of realistic impersonations, which were often limited to public figures, cases that were clearly criminal, or situations that were sufficiently serious and obvious enough to be addressed through existing law.

12. However, AI has changed this. It has made convincing, low-cost digital replicas easy to create and disseminate at scale without specialist or technical knowledge. Ordinary people are now frequently imitated in non-consensual ways that undermine their dignity and reputation resulting in financial and emotional costs, but which do not always meet the bar for criminal, civil or regulatory action. The government has already acted in respect of the worst cases, by strengthening criminal protections for AI-generated IIA and CSAM. However, there are many other non-consensual replicas that can fall outside of those offences or where there may be a need to go further.
13. Existing protections do not give most individuals meaningful control of their image or voice or access to realistic remedies. Creative stakeholders, for example, have told us that the UK's Intellectual Property (IP) framework is insufficient to protect against unauthorised imitations impacting their livelihoods.
14. There is, therefore, a potential case for giving individuals - including public figures, creative professionals, and the general public - better control over how their likeness, voice or personality can and cannot be digitally replicated. The government proposes to explore whether there are further circumstances in which unacceptable imitation, across sectors and people's day-to-day lives, should be prohibited. This will include exploring the case for greater commercial protections, and whether these should form part of wider safeguards or rights to personal image within the UK. The government's consultation on AI and Copyright included some initial questions on this issue, and the specific commercial-creative impacts of digital replicas are set out below.

Commercial-creative impacts of digital replicas

15. Digital replicas are a key area of concern to the creative industries. Digital replicas can take many forms, including music tracks, music videos, and film and TV content imitating the voice or appearance of real artists and performers. Digital replicas might be made and disseminated by private individuals on the open internet, including on social media and music streaming platforms, or they might be made and used in a commercial setting, for example by a record label or TV production company. When the technology is used commercially, and with the consent of individuals, it can have many benefits. It can provide new revenue opportunities for performers and creators, by allowing them to create and monetise more content. It can also be used to create content and techniques that would have been impossible or prohibitively expensive to achieve otherwise.
16. For example, a musician could use their digital replica to produce new music, increase fan engagement, or exploit a wider range of commercial opportunities. A film studio could artificially generate a large crowd in a film, de-age an older actor, or include a deceased actor in a new film. Recently, a film studio was able to de-age Harrison Ford in Indiana Jones and the Dial of Destiny to tell the story in a more convincing way. The technology can also be used to make existing production techniques cheaper and more effective. These uses can offer cost savings for content producers, new revenue streams for performers, and improved creative content for audiences. However, where digital replicas are used in place of performers and creators, this often comes at a cost to them.
17. Where digital replicas are made without consent of the individuals involved there may be negative commercial impacts. The digital replica can compete with the real

performer or creator's work, diverting revenue away from them and affecting employment opportunities. There is also the risk the consumer is deceived or confused, resulting in false endorsement, reputational or professional harm for the individual. One example is the 'fake Drake' track Heart On My Sleeve released on streaming platforms in April 2023. The track featured AI-generated vocals mimicking artists Drake and The Weeknd, and it amassed over 9 million views and streams before it was taken down. In another example, Stephen Fry's voice was recreated using AI to narrate a documentary without his permission. There are also examples of actors having their digital replicas used in ways they did not consent to.

Existing UK intellectual property law

18. There is no specific intellectual property protection relating to digital replicas in UK law, but existing rights will be relevant in some cases. IP rights may be applicable if an individual wants to commercialise elements of their personal likeness, or seeks redress for their unauthorised use, but there are gaps in what is protected. In general, intellectual property law does not protect ideas or concepts, it protects the expression of those ideas. As such, it may not be suitable for protecting a person's characteristics in general but may be useful in protecting a specific fixation of it. For example, it is unlikely that current intellectual property laws could protect a person's voice per se, but it may be applicable if a performance of a distinctive catchphrase is copied. The common law tort of passing off may be useful in some cases if the individual has "business good will or reputation", but this will not always be the case.
19. A short overview of the relevant intellectual property law is outlined below.

Copyright law

20. Copyright law protects works in which an individual's voice or appearance may be contained (such as sound recordings, films and photographs), but the protection applies to the work itself rather than to the identity of the person within the work. For example, in the case of a song, copyright protects the sound recording itself (section 5A Copyright, Designs and Patents Act 1988) alongside the lyrics (as a literary work – section 3 CDPA) and the music (as a musical work – section 3 CDPA), but it does not protect the voice or vocal characteristics of the singer.
21. Copyright is intended to protect the fixed expression of a creative work rather than the underlying ideas or stylistic characteristics. Certain acts will infringe the copyright in a work if a "substantial part" (or the whole) of the work is used – for example, if a substantial part is copied. Therefore, if an AI model is trained on copyright works to produce a digital replica, the digital replica itself will infringe copyright only if it reproduces a substantial part of an existing copyright work, notwithstanding potential infringement during AI model training. A digital replica will not necessarily infringe copyright if it imitates a person's face or voice, but there could be a potential infringement if it is a copy of the whole or a substantial part of an existing copyright work.
22. Even if a digital replica infringes the copyright in an existing work, the person whose identity was compromised in the digital replica is unlikely to be the copyright owner. For example, in the case of sound recordings it is the producer who first owns the copyright under the CDPA. In the case of films, it is the producer and the principal

director who first own the copyright under the CDPA. In the case of photographs, it is the photographer rather than the subject of the photo who first owns the copyright under the CDPA. Therefore, the person whose voice or image had been copied is unlikely to have direct access to recourse via copyright law. Whilst a recording artist enjoys rights to control the recording, copying and distribution of their performance, we have anecdotal evidence that, in practice, these rights are often transferred by licence or assignment to the producer or director. Therefore, whilst the recording artist may benefit from action taken by the copyright owner, they are unlikely to be able to initiate this action themselves.

Performers' rights

23. Performers' rights are a type of related right, provided for in Part 2 CDPA. Under section 180(2), 'performance' means (a) a dramatic performance (which includes dance and mime), (b) a musical performance, (c) a reading or recitation of a literary work, or (d) a performance of a variety act or any similar presentation, which is (or so far as it is) a live performance given by one or more individuals. A 'recording' of that performance means a film or sound recording (a) made directly from the live performance, (b) made from a broadcast of the performance, or (c) made, directly or indirectly, from another recording of the performance.
24. Performers' rights confer economic and moral rights on performers with respect to their performances. The rights last for 50 years from the end of the calendar year in which the performance took place. If, during those 50 years, a recording (other than a sound recording) of the performance is released, the rights last for 50 years from the end of the calendar year in which the recording was released (and in the case of sound recordings, this is extended to 70 years).
25. Performers' economic rights confer both an 'ability to prohibit' and an 'ability to authorise'. This is achieved through the way in which these rights are framed as both property and non-property rights which are infringed unless the performer consents to the act in question.
26. The economic rights can be broadly categorised into three sets:
 - The right to consent to the making of a recording (or broadcast) of a live performance (section 182)
 - The right to control the subsequent use of such recordings, such as right to consent to a copy of the recording being made or being distributed to the public (sections 182A – 184)
 - The right to receive equitable remuneration for the exploitation of sound recordings (section 182D). However, this economic right is not relevant to the issues surrounding unauthorised digital replicas.
27. Performers' rights are tied to the recording of a performance, not the substance, style and content of the performance itself. If the digital replica does not contain a recording in which performers rights subsist, then they cannot be relied upon by the performer. For example, if AI is used to generate a new performance which includes a digital replica of a performer, it is unlikely they would be able to rely on performers' rights, as the protected performance is not present in the output.

28. Including performances in the sphere of copyright draws on the thinking of Lord Justice Arnold who, in his book *Performers' Rights*,⁸³ argues: 'The first task is to bring performers into the copyright fold proper, rather than to continue to pretend that performers' rights are in some way different to other copyrights'. However, Lord Justice Arnold does note there are some outstanding questions if copyright was extended to performers given performer's rights are not infringed by copying or a reproduction of a performance itself⁸⁴.

Moral rights

29. In addition to economic rights, authors, film directors and performers are granted moral rights under the CDPA. These are known as the 'attribution right' and the 'integrity right'. The attribution right refers to the right to be identified as author or director (section 77) or performer (section 205C). This right must be asserted in writing in order for it to be enforced. Authors and directors have an additional right to prevent false attribution (section 84). The integrity right refers to the right to object to derogatory treatment of a work (section 80) or performance (section 205F). 'Derogatory treatment' means any distortion or mutilation of the work / performance or any other modification that is prejudicial to the reputation of the author, director or performer. Moral rights are not transferable, but they are waivable.

Trade marks

30. Trade marks are 'badges of origin' identifying the source of products or services. To perform this function, they must be distinctive. Trade marks are most commonly used in the commercial world to distinguish and protect brands and businesses.
31. Under UK trade mark law, a person can theoretically register their name, signature, nickname, sound clips of their voice, or images or videos of themselves as a trade mark. If a person successfully registers an aspect of their identity as a UK trade mark, they may be able to cite trade mark infringement if the trade mark is used in a digital replica without their consent. It would have to be proven that the digital replica had commercial motivations or implications, since trade mark infringement can be invoked only where it is used 'in the course of trade'.
32. Although a sound clip of a voice can be registered as a trade mark, this may prove challenging in practice due to the distinctiveness requirement in UK law. An individual's voice per se cannot be registered as a sound mark. Sounds can only be registered if they are 'capable of distinguishing the goods or services of one undertaking from those of other undertakings'⁸⁵. In other words, the sound must be exclusively associated with one business undertaking for the goods or services it is registered for. If a clip of a voice is successfully registered as a trade mark, the scope of protection will depend on the specific words spoken in that recording. As a result, it may be difficult to rely on the mark to take action against a digital replica that uses a similar-sounding voice to say different words.

83 Richard Arnold, *Performers Rights* Sixth Edition, Sweet & Maxwell, Proposals for reform para 1-116. ISBN: 978-0-414-0809

84 Ibid para 1-123

85 See Manual of Trade Marks Practice: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/361114/Manual_of_trade_marks_practice.pdf

33. Regarding trade marks for images or videos of a person, the scope of protection offered will again depend on what is contained within those particular images or videos and whether infringing use is identical or confusingly similar to the registered trade mark. Such registrations would not offer general protection against the use of a person's image in a digital replica without their permission.

Passing off

34. Passing off is a common law tort. It is intended to prevent businesses misrepresenting ('passing off') their goods or services as the goods or services of another business. To bring a claim of passing off, three elements are required,
- the existence of business goodwill and/or reputation on the part of the claimant;
 - a misrepresentation by the defendant which is likely to mislead consumers as to the origin of the goods or services;
 - damage to the claimant as a result of the misrepresentation.
35. Passing off typically relates to goods and services in a business-to-business context, although case law⁸⁶ has shown the tort can be applied flexibly depending on the specific circumstances of a case. It is possible that the tort could be applied to a commercially available digital replica of a famous person, which another individual or organisation 'passes off' as authentic content by that person. For a claim of passing off to be successful, the three criteria (goodwill and/or reputation, misrepresentation, and damage) would need to be fulfilled.
36. The applicability of passing off to unauthorised digital replicas has not been tested in the courts yet. While popular artists may be successful in a claim of passing off, it is unlikely that lesser-known artists would have the level of reputation required to make a successful claim. It also would not address the personal identity/moral rights concerns of stakeholders. It is also unclear if the requirement of misrepresentation would be met if an unauthorised song was labelled as AI-generated.

International law and developments

37. Protection for personality exists in different forms in other jurisdictions.
38. Several EU countries have well-established personality rights frameworks. These tend to focus on privacy and moral aspects, rather than commercial interests. For example, Germany, Italy and the Netherlands include image rights within their copyright legislation. Alongside image rights, Italy has developed a strong, commercially-focused right of publicity.
39. More recently, Denmark and the Netherlands have put forward legislative proposals. Denmark has proposed introducing a performer-specific digital replica right to their neighbouring rights law. The proposal would enable performers and artists to control the making available of digital replicas of their performances. The Danish

⁸⁶ See *Irvine v. Talksport Ltd.* (2002) and *Robyn Rihanna Fenty v. Arcadia Group Brands Ltd* (2013)

government is also proposing to introduce a digital replica right for the general public within its neighbouring rights framework. The Netherlands is also proposing to introduce a digital replica right in its Neighbouring Rights Act that applies to the general public.

40. In the USA there is no federal personality right. However, many USA states provide rights of publicity. These are property rights which are designed to prevent the misappropriation of an individual's name, likeness, or other identifying feature for commercial purposes. The right exists in most USA states, either by statute or common law. Since it is not a federal right, its characteristics and levels of protection differ from state to state, but in all cases the right can be said to protect the value in one's identity.
41. More recently there have been some legislative developments to specifically address the issue of unauthorised AI-generated digital replicas. In March 2024, Tennessee passed a new law titled the 'ELVIS' Act (Ensuring Likeness Voice and Image Security Act) which amends and extends the state's publicity rights framework to account for AI-generated content. We also note the California Assembly Bills 2602 and 1836, which provide protection to performers in the context of AI-generated digital replicas. At a federal level, the 'NO FAKES' Act (Nurture Originals, Foster Art, and Keep Entertainment Safe Act), intends to tackle the misappropriation of image, voice and visual likeness in 'digital replicas'. The 'No AI FRAUD' Act (No Artificial Intelligence Fake Replicas and Unauthorized Duplications Act), introduced in the House of Representatives in January 2024, is similar to the NO FAKES Act and intends to 'provide for individual property rights in likeness and voice'.
42. More work is needed to understand these, and other, international approaches, as we consider our next steps in this important area.

Views from the consultation

43. The Copyright and AI Consultation included two questions seeking general views on digital replicas. It asked to what extent the Government's proposals on copyright and input transparency would provide sufficient control over the use of image and voice in AI outputs and sought people's experience and evidence on digital replicas (questions 43 and 44).
44. In answering question 43, the most common theme identified was that the measures outlined in the first part of the consultation would not provide sufficient control. A less common theme identified was that further legislation was needed. In response to question 44, respondents provided a range of experiences of digital replicas. Many of these focused on the impacts digital replicas could have on performers and creators. Many responses pointed to the issue of outputs being generated that replicate creative works. Some respondents pointed to the malicious, illegal and/or harmful content that AI outputs can generate, including issues such as impersonation, defamation, sexual harassment, fraud and misinformation.
45. Most responses to question 43 shared the view that the approaches outlined in the first part of the consultation, in relation to transparency and a text and data mining exception, would not provide individuals with sufficient control over the use of their image and voice in AI outputs. Some respondents noted that transparency measures over both inputs and outputs are important and may help individuals

control their voice and likeness to some extent, but transparency measures alone would not prevent unauthorised use.

46. Many responses to both questions 43 and 44 argued that greater legal protections for image and voice were necessary. There was no single view on what form this additional protection should take, with respondents advocating for several different interventions depending on how they are impacted by digital replicas. For example, some performers called for interventions that would give them unassignable control of their voice and likeness, whereas organisations representing those performers commercially often said any new rights should be assignable.
47. Some respondents outlined the general limitations of copyright law in providing protection against unauthorised digital replicas. These include the fact that copyright protects creative works rather than a person's characteristics (even if those characteristics are embodied in a work), and the difficulties of enforcing existing legislation. For example, record labels Warner Music Group and Sony Music reported that several thousands of unauthorised voice clones of artists they represent have been uploaded to music streaming services in recent years. The labels said they have issued takedown notices, with mixed success. The British Film Institute also highlighted the importance of making sure existing law is properly enforced.
48. Issues with contract were raised by several respondents. While noting that AI and digital replication clauses are becoming more prevalent in contracts, some creators said that contractual provisions are often inadequate or unclear, and do not provide the artist with sufficient consent, remuneration or transparency over the use of their voice and image. Some respondents, like Equity, raised concerns around contracts with unfavourable terms for performers, particularly those negotiated before the AI era, being relied upon in the context of new technologies.
49. Several respondents, particularly within the creative industries, advocated for the implementation of legislative measures. There was no single view on what form the intervention should take, its scope or who should be liable when it is infringed. Suggested interventions included creating personality rights and specific digital replica rights and amending performers rights.
50. For example, Equity, the UK's performing arts and entertainment trade union, called for a new suite of IP rights for performers, including new image rights and updated performers' rights and moral rights, to address specific issues performers are facing with an increase of AI-generated performances. Others, like the Council of Music Makers, which represents a range of music creators, supported a new digital replica right. They proposed that a new right could help address concerns that unauthorised AI digital replicas may result in misappropriation and false endorsement. A new right could also provide opportunities for the commercialisation of digital replicas. Other respondents proposed the broader intervention of introducing personality rights into UK law. Some respondents argued that users, rather than AI developers, should be liable if AI tools are used to generate unauthorised digital replicas.
51. There was support for further consultation before legislative intervention. These respondents emphasised that any legislative interventions must be proportionate and carefully considered – and may extend beyond the scope of copyright law. The British Copyright Council, an organisation representing all parts of the creative industries, highlighted that the impact of any new measures would need to be

considered thoroughly and holistically: ‘Careful analysis and assessment of what changes to the legal framework would mean for the application of longstanding industrial relations procedures within creative industry production sectors is needed, before recommendations for change are put forward.’ Organisations including AI firm Anthropic and trade body TechUK emphasised that any approach to regulating digital replicas should be balanced and tailored. Anthropic, for example, said that any new rules should be carefully crafted to ‘avoid overbreadth that impedes new creativity and expression’.

52. Some respondents suggested that the government should undertake further research on the impact of digital replicas as a first step. For example, Creative UK proposed that the Government should research the impact of digital replicas on creators to further understand how current law provides protection over voice and appearance, and where new legislation might be required. Dr Mathilde Pavis, an academic, lawyer and consultant specialising in IP, AI and digital cloning, proposed that the government should undertake an evidence-gathering process to map out the emerging ecosystem of the ‘digital replica economy’. This could then inform any necessary regulatory interventions. Some technology companies recommended the government research claims or complaints that are being handled under existing laws and use those findings to explore protections via existing and more appropriate channels such as online safety, privacy and criminal law.
53. Several respondents across all sectors noted that digital replicas span many areas of law, including privacy, data protection, licensing and tort law. They advised the Government to explore options beyond IP and look at international examples. For example, the record label sector was supportive of the proposed USA ‘NO FAKES Act’ which would introduce a federal right for individuals to control the use of their voice and likeness in digital replicas.
54. Some respondents commented that data protection law should be able to provide remedies for the unauthorised use of voice and likeness, but it was difficult to enforce. For example, Equity said that performers’ data protection law ‘is not being properly enforced through the courts’. The British Equity Collecting Society, representing audiovisual performers, said that the Information Commissioners’ Office (ICO) needs to be ‘adequately resourced’ to take action against breaches of data protection law. The Oxford Intellectual Property Research Centre analysed the benefits and shortcomings of data protection law in providing protection for unauthorised digital replicas, while the ICO said that ‘a combination of different approaches may need to be considered’.
55. Not all respondents thought legislative intervention was necessary. Producers from the TV and film industry and the tech industry generally thought the existing law provides sufficient control over voice and image in AI outputs, as long as that law could be enforced properly, and that any intervention risked unintended consequences. For example, the Motion Picture Association (the American association of film and TV producers) cautioned against new legislation, saying ‘Any legislation in this area, even if properly limited to targeting unauthorised digital replicas, risks inadvertently chilling the fundamental right to freedom of expression, which is the essence of storytelling as well as parody.’ The Producers’ Alliance for Cinema and Television similarly cautioned against new legislation. In their view, robust transparency measures around AI model training were needed instead.

56. Some respondents, particularly authors and visual artists, were concerned about the ability of generative AI to imitate style via the input of ‘in the style of’ prompts. The Association of Photographers said that this allows a user to “...generate hundreds of mimicked references that a user can commercially exploit, whilst making them complicit in ‘data-laundering’ without the artist’s consent”.

Stakeholder engagement

57. The issue was discussed at the technical working group focussing on wider support for creators. Several members of the group viewed this as a complex issue, with a complicated legislative picture, and that different groups would be impacted differently.
58. Representatives from the creative industries spoke in favour of strengthening protections around voice and likeness by introducing a new right. Some outlined how digital replicas are displacing creators and performers work and earnings, with images and voices being used without consent, often from contracts signed before the current AI landscape. Existing power imbalances in contractual negotiations were also felt, by some, to be an issue. This power imbalance may result in performers and creators being unable to challenge how copies of their voice or likeness will be stored or used in future.
59. Those advocating for a new right generally felt the existing legal frameworks like IP and General Data Protection Regulation (GDPR) are insufficient, either because they are out of date or not suitable for the specific issues creators and performers are facing. The complexity of the licensing environment associated with creative content was also discussed, and some felt this complexity means it can be cheaper to create new content rather than to license existing material. Some advocated for a new right to address the issue of AI being used to create material in the style of well-known creators and performers. Others thought this a distinct issue to that of AI-generated deepfakes of individuals, involving different legal considerations.
60. The group also discussed existing initiatives to address the issue, from technical tools to detect unauthorised digital replicas, to the proposed legislative approach in the USA NO FAKES Act. Whilst the voluntary initiatives were acknowledged by the group, many felt legislation was necessary.
61. In complement with the technical working groups, the government has also met with stakeholders to discuss digital replicas in more detail. Representatives for creators, right holders, production company trade bodies and collecting societies were asked questions about four topics relating to digital replicas: financial and ethical impacts; contractual terms; current UK protection; and recommendations for government intervention.
62. Regarding financial and ethical impacts, most sectors viewed digital replicas as a threat, but there was also agreement that they can provide commercial opportunities, particularly for more famous actors, models and musicians, and their associated right holders. Additionally, the representatives of production companies pointed towards the financial benefits of utilising digital replicas and questioned whether this always results in job displacement. When considering contractual terms, although digital digital replicas are not explicitly addressed in UK legislation,

representatives indicated that creators are negotiating for the use and control of digital replicas in contracts with licensees and right holders.

63. Regarding current UK legislation, it was noted that passing off law, the Defamation Act, and the Online Safety Act, and GDPR have some potential to control digital replicas. However, aside from the employment of GDPR by some sectors, little use is being made of this legislation. Additionally, it was argued by all representatives of creators, right holders and trade bodies that current UK legislation is insufficient to fully address the protection and commercial exploitation of digital replicas. In respect of recommendations for government intervention, there was greatest support for a personality right. Some representatives argued that the protection for digital replicas should not form part of copyright law, as it is designed to protect expression rather than style.

Conclusion and proposals

64. AI makes it easier to create ‘digital replicas’ of someone’s voice or face. This can be a powerful tool, including for the creative industries. However, when someone’s likeness is replicated without their permission it can cause harm to the replicated individual and to others who consume the content, and can also constitute illegal content. While there are some protections in place today, these do not cover the full risks associated with digital replicas.
65. The responses to the consultation show this is an area of growing concern for the creative industries, especially for performers. There is clear concern that the existing legal framework may not be sufficiently robust to deal with unauthorised digital replicas. Many stakeholders support enhanced protections for a person’s image and voice. However, there was no single view on what form this should take or to whom they should apply.
66. As seen through the consultation and stakeholder engagement, many performers and creators think there is no adequate, reliable legal route to redress. While the more successful artists may be able to rely on passing off in some scenarios, this avenue is less accessible for lesser-known artists or the general public. Copyright law can only be relied upon when a substantial part of an existing work is copied. Subsequent enforcement action can only be taken by the copyright owner, and this may not be the subject of the digital replica. Performers rights are also hard to rely on as they are tied to a recording of an existing performance and will not be helpful if a new AI-generated performance is made.
67. There are areas of law beyond IP that the AI copyright consultation did not ask questions about. These include data protection and privacy law. Again, where respondents commented on these, there was no consensus that they provide a more suitable legal framework to deal with digital replicas.
68. We agree that the growing use of realistic impersonation through AI creates new risks – as well as opportunities – for artists and the general public. We propose to explore options that address these risks, while promoting growth and innovation. This will include considering whether a new personality right may be appropriate.

