OpenAI

UK AI and Copyright Consultation Summary
2 April 2025

OpenAI is a leading artificial intelligence research and deployment company. Our mission is to ensure that artificial general intelligence — in short, AI that's at least as smart as a person — benefits everyone. In service of our mission, we research, develop, and release cutting-edge AI technology as well as tools and best practices for the safety, alignment, and governance of AI. Our goal is to build AI systems that help solve hard problems in areas like healthcare, science and education to contribute to a world where everyone benefits from the social, economic, and technological opportunities of AI.

We believe the most effective way for the Government to achieve its stated objectives of unlocking and leveraging the broad benefits of AI in the UK is through a broad text and data mining exception, as described in Option Two of the consultation proposal. Our position is based on three key principles:

1. **The UK has a robust AI ecosystem, but global competitiveness requires treating data access as a fundamental building block for AI-driven economic growth and investment.** The UK can become a global leader in AI by adopting policies that build on its current strengths. British researchers have been instrumental in advancing the AI field, and today the UK is one of the leading countries in the world for AI adoption. However, seizing this potential will require a policy framework that recognizes the crucial role of data in developing advanced AI systems, creating new AI-driven applications to solve hard problems, and driving economic growth and investment.

2. **A broad TDM exception is necessary to drive AI innovation and investment in the UK, and can be made to balance the needs of AI development with the mitigation of concrete harms to copyright owners.** Innovation and investments in AI infrastructure are only possible in jurisdictions where laws clearly support technological research and development. The UK must create a clear, predictable regulatory environment that sets it apart from other jurisdictions in order to boost its competitiveness. Such laws are not zero-sum: the Government can implement a broad TDM exception (Option Two) in a way that encourages innovation, addresses specific rightsholder concerns, and recognizes that AI will create new opportunities for creators' and rightsholders' creative output.

3. **The EU regime shows the challenges of implementing opt-out systems.** In the EU, the lack of clear and scalable technical standards has created uncertainty about what opt-out methods are workable and valid, causing uncertainty for both AI companies and rightsholders. In contrast, the US approach has sustained American leadership in earlier technological waves and now AI. The UK has a rare opportunity to cement itself as the AI capital of Europe by making choices that avoid policy uncertainty, foster innovation, and drive economic growth.

Our full response to the AI and Copyright consultation is below. We look forward to working with governments and creators everywhere to shape the future of AI innovation and copyright.

OpenAI

UK AI and Copyright Consultation Response
Submitted via IPO survey, 25 February 2025

Question 4. Do you agree that option 3 [Option 3: A data mining exception which allows right holders to reserve their rights, underpinned by supporting measures on transparency] is most likely to meet the objectives set out above?

OpenAI is a leading artificial intelligence research and deployment company. Our mission is to ensure that artificial general intelligence — in short, AI that's at least as smart as a person — benefits everyone. In service of our mission, we research, develop, and release cutting-edge AI technology as well as tools and best practices for the safety, alignment, and governance of AI. Our goal is to build AI systems that help solve hard problems in areas like healthcare, science and education to contribute to a world where everyone benefits from the social, economic, and technological opportunities of AI.

We believe a broad text and data mining exception, as described in Option 2, is the most effective way to achieve the Government's objectives to unlock and leverage the benefits of AI in the UK. Our position is based on three key points: 1) the UK has a robust AI ecosystem, but global leadership requires treating data access as a fundamental building block for AI-driven economic growth and investment; 2) Option 2 will better balance the Government's objectives than Option 3, as the latter risks negative consequences for AI uses that would otherwise benefit the UK broadly; 3) the EU regime is instructive of significant implementation challenges with Option 3.

1). UK leadership in AI is contingent on adopting a policy framework that recognizes the crucial role of data.

We believe that the UK can become a global leader in AI by adopting policies that build on its current strengths. British researchers have been instrumental in advancing the AI field, and today the UK is one of the leading countries in the world for AI adoption. The AI Opportunities Action Plan will further strengthen the country's AI ecosystem through infrastructure investments, piloting promising applications of AI and investing in domestic champions. We're also excited to see the UK Government supporting small businesses to quickly find tailored advice on GOV.UK by testing a chatbot built using OpenAI's GPT-4o model.

However, seizing all of this potential will require a policy framework that recognizes the crucial role of data in developing advanced AI systems and creating new AI-driven applications. Data access is an essential building block for AI-driven economic growth and competitiveness, which will secure the UK's place among other countries leading AI research and development. Such a framework will also ensure the UK can both shape and leverage the benefits of AI, including improved public services in areas such as healthcare and education, and increased productivity to address much-needed economic growth.

2). Option 3 does not evenly balance the stated objectives of this Consultation.

We are skeptical that Option 3 will achieve the Government's three objectives of 1) providing copyright owners control over use of their works, 2) providing easy access to AI developers to sufficiently large volumes of data, and 3) building a transparent and sensible copyright framework for AI development. Option 3 prioritizes the first objective over the other two by giving content owners a degree of control that outweighs the other objectives. If, as the Government proposes, reserving rights against any text and data mining can be done collectively and "easily", there is a risk that most copyrighted data will become unavailable for text and data mining uses. Under such a system, the Government's access objective would not be met, because only the wealthiest and already data-rich technology companies would be able to access the quantity and diversity of data needed to train advanced models. This would not be a workable or sensible system for AI developers, who would be forced to disclose their most sensitive training details in order to strengthen the negotiating position of copyright owners. This situation would be similar in practice to Option 0, except with added regulatory costs on AI developers to comply with transparency and other regulatory measures the Government adopts to the detriment of AI development.

Additionally, an opt-out approach will block AI applications that do not compete with or diminish the value of human-created works, obstructing whole sectors of future economic growth and stifling the UK's ambitions at technological leadership. For example, many AI-powered tools — ranging from language translation and spam detection to medical research and climate modeling — require large-scale data analysis to function effectively. AI will enable broad economic growth and shape the competitive global landscape, so the UK should view its own leadership in terms of seizing the growth potential from AI and ensuring democratic governments are leading AI development.

Restricting the data on which these technologies can train will hinder advancements that would benefit society as a whole, including rightsholders, and may have undesired consequences for the UK's economic and national security objectives. Instead, as discussed in our response to Question 5, we believe that Option 2 can be constructed to balance the interests of content owners with the needs of AI developers.

3). The EU experience with an Option 3 approach reveals significant implementation challenges.

As the Government notes, the opt-out framework in the European Union has significant implementation challenges that undermine all three objectives. The lack of clear technical standards has created uncertainty about what opt-out methods are valid, leading to considerable variance in how rightsholders attempt to reserve their rights. As a result, AI developers struggle to identify which works can be accessed and which are off-limits, a task that becomes unmanageable at scale without machine-readable protocols. Furthermore, the EU transparency requirement is currently being drafted, so it remains unclear how the obligation will operate in practice and whether it will both satisfy rightsholders and be workable for AI developers. It's important to also note that the EU transparency requirements are sought primarily by rightsholders to increase licensing leverage and initiate litigation against AI developers, which risks creating a chilling effect on AI development in the EU. Should the UK follow this approach, it could lead to a significant setback in achieving the Government's AI-driven growth goals. Instead, the UK has a rare opportunity to cement itself as the AI capital of Europe by making choices that avoid the policy uncertainty currently restricting AI investment in the European Union.

Question 5. Which option do you prefer and why?

- Option 0: Copyright and related laws remain as they are
- Option 1: Strengthen copyright requiring licensing in all cases
- Option 2: A broad data mining exception
- Option 3: A data mining exception which allows right holders to reserve their rights, supported by transparency measures

We believe the Government should more seriously consider a broad TDM exception as described in Option 2. To become a global leader in AI, the UK must establish a clear, predictable regulatory environment that sets the UK apart from other jurisdictions like the EU. Importantly, such an exception is not zero-sum: the Government can implement Option 2 in a way that encourages innovation, addresses specific rightsholder concerns, and recognizes that AI will create new opportunities for creative output.

1). A broad TDM exception is necessary to drive AI investment in the UK.

The decisions the UK makes today will shape the future of its AI sector. AI developers are actively investing tens of billions of pounds in physical infrastructure as they plan for future growth. This includes OpenAI, which recently announced the Stargate Project to invest $500 billion in building AI infrastructure in the United States (https://openai.com/index/announcing-the-stargate-project/). Large infrastructure projects are long-term commitments of capital and people, so companies invest in jurisdictions where laws clearly support technological research and development. In our case, investing in the United States was possible 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.

By contrast, narrow or complex exceptions will not help the UK in attracting AI development. An opt-out framework under Option 3 would suffer from technical complexity and require a lengthy multi-stakeholder process to develop effective technical solutions. A British AI researcher who wants to start a company will consider laws in different jurisdictions and the costs of operating in them. If the UK does not meaningfully differentiate itself from jurisdictions such as the EU, which have larger markets, it will not firmly establish itself as one of the leaders in the global race for attracting AI businesses, infrastructure investments, and talent.

2). The Government can design Option 2 to balance the needs of AI development in the UK with the mitigation of concrete harms to copyright owners.

Rather than designing Option 2 with “no restrictions”, the Government can condition a broad TDM exception based on three requirements: 1) developers should use lawfully obtained copies; 2) copies used for training are not further distributed or used for other purposes; and 3) developers should take reasonable measures to prevent infringing outputs.

The first requirement would address the concern raised by some copyright owners that AI developers may use pirated materials to train their models. Both the European Union and Singapore require use of lawfully accessed copies: Article 4(1) of the 2019 EU Copyright Directive permits text and data mining

only of “lawfully accessible works”, and Section 244(2)(d) of Singapore’s Copyright Act requires beneficiaries to have “lawful access to the material” being used. A similar requirement in the UK would ensure that AI developers obtain the works they use for training through legal channels, allowing rightsholders to receive compensation.

For the second requirement, Singapore could serve as a useful example, as it requires that copies used for text and data mining are “not use[d] for any other purpose” (Sec. 244(2)(b)). For the third requirement, because copyright owners are most directly harmed by infringing AI outputs, the Government could require AI developers to take reasonable measures to prevent them, though such measures would need to account for lawful outputs, as we discuss further in our answer to Question 44.

Crucially, a broad TDM exception does not mean that rightsholders will receive no remuneration from uses of their works. Even in jurisdictions with broad copyright exceptions for text and data mining, AI companies strike partnerships with content owners, such as to access works that are not available on the open market (such as archival materials or B-roll footage). Companies also partner with copyright owners to unlock specific applications of AI in products that require close cooperation and expertise in both technological and creative fields, such as OpenAI’s partnerships with news organizations to facilitate deeper forms of display and interaction with current news content in search. Though the industry is still nascent, we see signs that there will be more partnerships between AI developers and creatives. For example, the startup Rebind partners with authors to let users deeply engage with classic books through expert commentary, videos, and AI-driven interactive discussions.

In addition, we share concerns about the use of AI tools to create unauthorized digital replicas, and we have supported legislation prohibiting the creation and distribution of such replicas, such as the NO FAKES Act in the United States. If the Government enacted a broad text and data mining exception, it could be paired with a law prohibiting the use of AI or other digital technologies to create unauthorized replicas of artists’ likeness or voice. This would allow the Government to address harmful uses of AI technologies without harming uses that benefit society.

Training is a foundational step to building AI systems, and legal restrictions on training will have outsized effects on the technology as a whole. AI outputs, on the other hand, particularly those that replicate copyright-protected expression or imitate an artist’s likeness or voice, raise clear concerns for copyright owners and artists. Therefore, the Government should focus its regulatory interventions on AI outputs and preventing users from evading model restrictions to intentionally generate infringing outputs, rather than restricting inputs that are used for model training.

3). AI is a tool that creates, rather than replaces, opportunities for creative output.

AI is already proving itself as a tool that expands creative output and enables new forms of artistic expression. As we’ve seen through our partnerships with creators, AI enhances human creativity by providing new ways to explore ideas, experiment with different styles, and push the boundaries of artistic expression. Rather than limiting originality, it serves as a collaborator that expands the creative process, allowing artists to bring their visions to life in ways that were previously impossible or impractical. For example, artists have used our video generation model, Sora, as part of a broader creative workflow to

create new types of video, some of which were showcased at the Tribeca Film Festival (https://www.nbcnewyork.com/entertainment/the-scene/tribeca-festival/openai-sora-shorts-debut-tribeca-film-festival/5515224/ and https://tribecafilm.com/films/sora-shorts-2024).

International experience also clearly demonstrates that well-designed TDM exceptions and more flexible copyright regimes do not cause economic losses to the creative industries. The US is a good example for a jurisdiction that has both a thriving AI sector and creative sector – with the latter continuing to grow despite AI applications having gone mainstream (see, for example, CCIA's 'The Sky Is Rising' 2024 report: https://ccianet.org/research/reports/sky-is-rising-2024-edition/#main-content). Similarly, since Japan adopted its 2019 copyright exception for uses that do not enjoy a work's thoughts or sentiments, its music and animation industries have continued to grow (see, for example, International Federation of the Phonographic Industry (IFPI)'s Global Music Report for 2024 https://globalmusicreport.ifpi.org/ and the Association of Japanese Animations' Anime Industry Report 2023 https://aja.gr.jp/download/2023_anime_ind_rpt_summary_en). We are also not aware of evidence of increased copyright infringement in these jurisdictions stemming from the availability or use of AI tools.

Question 6. Do you support the introduction of an exception along the lines outlined above?

See responses to Questions 4 and 5.

Question 7. If so, what aspects do you consider to be the most important?

See responses to Questions 4 and 5.

Question 8. If not, what other approach do you propose and how would that achieve the intended balance of objectives?

See responses to Questions 4 and 5.

Question 9. What influence, positive or negative, would the introduction of an exception along these lines have on you or your organisation? Please provide quantitative information where possible.

First, the introduction of an exception based on Option 3 would undermine the UK's talent advantage, which has driven the UK's contributions to fundamental AI research. As discussed in our earlier responses, investment into AI infrastructure will flow to jurisdictions with regulatory environments that encourage AI development. Talent will concentrate where AI development and model training take place. Because Option 3 will impose more restrictions on AI development than other copyright regimes, AI researchers will likely flow from the UK into other countries, a direct and negative effect that UK AI startups will feel most acutely.

OpenAI will also be affected by changes to the UK talent pool. The UK was our first choice for opening an international office in 2023, in part because it gave us access to the UK's research and engineering talent. While we currently have a small research and engineering staff in the UK, regulatory changes that reduce

the UK's competitive edge compared to other jurisdictions will make it difficult for us to expand our research footprint in the UK. By contrast, a more forward-looking and innovation-enabling copyright regime would attract and cultivate more research talent to the UK from other jurisdictions, giving us more options when hiring exceptional talent.

Option 3 would impose significant additional costs and compliance burdens across the AI development lifecycle on OpenAI and other developers. Because AI models are frequently adjusted or fine-tuned, 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. This would increase compliance costs, slow model updates, and may result in the UK having access to less capable models than its international peers. This will have downstream effects on the UK's economic competitiveness, as its industries will not be able to realize the same benefits from access to the most advanced models.

Finally, the Option 3 approach is likely to result in further economic concentration in the technology sector, to the detriment of startups and medium-sized challenger companies like OpenAI. A licensing market for AI development will emerge as an extension of the current technology market, in which a very small number of large global firms generate vast profits and sit on large stores of data. A regulatory regime requiring AI training data to be done under copyright license will benefit the largest firms, who have large data stores to use in development and even larger profits to license whatever additional data they may need. This is particularly true given that the largest technology companies also own copyrights from their past expansions into creative sectors, including video game development, prestige television studios, and augmented reality. These are valuable sources of data that the largest AI developers will not need to license and would be unwilling to license to competitors.

Question 10. What action should a developer take when a reservation has been applied to a copy of a work?

Currently, it is not possible to make an AI model "forget" the patterns it has learned from seeing a piece of data during the training process. AI models do not store training materials in a database; they are software that makes predictions based on patterns they observed across an immense quantity of data. Because AI models do not store works, reservations should only apply to future model training.

Question 11. What should be the legal consequences if a reservation is ignored?

While we believe Option 2 is a better approach, if the Government adopts a version of Option 3, the established legal framework for copyright should govern a deliberate and repeated disregard of rights reservations in future model training.

However, legal consequences should account for inadvertent oversights, as well as the compliance challenges outlined in responses to Questions 4 and 5. Until there is a clear, effective technical solution for rights reservation, AI developers are likely to make mistakes and oversights. Developers should be given the opportunity to address any unintentional omissions, with legal measures only coming into play if, after being given a reasonable opportunity to rectify the issue, an AI developer continues to not apply

the reservation to future model training.

Question 12. Do you agree that rights should be reserved in machine-readable formats? Where possible, please indicate what you anticipate the cost of introducing and/or complying with a rights reservation in machine-readable format would be.

Yes, rights reservations should be machine-readable, and specifically in a form that can be automated. Other forms of rights reservation will be too complex for any company to administer and particularly harmful to startups. Because AI models are trained on massive datasets, a manual system would be highly complex and expensive to implement. Without a standardised approach, ambiguous or inconsistent opt-out signals could lead to a burdensome, manual process for both right holders and developers. Additionally, rights holders would benefit from a "one stop shop" whereby they use one signal for all developers, rather than having to adopt a different approach for each developer.

However, because there is no current machine-readable standard working at scale for AI training — especially at the level of individual works — this approach would require substantial investment and coordination across the creative and technology sectors. Right holders would need to adopt consistent metadata or tagging practices, potentially requiring changes to how they label and distribute content. Developers, in turn, would need to build infrastructure to ingest, update, and respect these opt-outs across multiple formats and platforms, adding further complexity and cost. There would also need to be safeguards to ensure that individuals own the rights they claim to own and that developers are taking the steps they claim to have taken to apply reservations.

Question 13. Is there a need for greater standardisation of rights reservation protocols?

The private sector is already working on standards focused specifically on AI training and rights reservation. Industry and standards-setting groups, such as the W3C, IETF, and IPTC, are actively discussing new standards, but those conversations have shown the difficulty of achieving consensus on common protocols and formats. OpenAI has and will continue to join these conversations, but it is important to remember that standards can take years to develop. Once that process is complete, standards endure because they are the result of collaboration and agreement among diverse stakeholders to ensure that every angle is considered and addressed.

With that said, we agree that there is room to improve the current state of rights reservations standards. In web crawling, for example, AI developers generally follow standards like robots.txt, but robots.txt was designed for web crawling and requires web publishers to either know and specify the name of a specific crawler or to set blanket rules for all unknown crawlers, regardless of their purpose. Moreover, standards like robots.txt that apply at the website level are not granular enough to differentiate between different files on a website, and are not a solution for handling analog content.

While we have seen interest in metadata-based approaches, such as from the Coalition for Content Provenance and Authenticity (C2PA), they too have limitations. Metadata can be stripped when content is rehosted, reformatted, or shared across different platforms, making it an unreliable mechanism for

large-scale rights management. Because metadata may be inconsistent across files, developers would need a way to determine the canonical metadata to apply to all representations of a particular work so that each version is handled consistently. Additionally, metadata is a forward-looking solution: it must be applied to works after a standard is adopted, but no copies of the works that exist today will bear metadata.

Question 14. How can compliance with standards be encouraged?

- No answer

Question 15. Should the government have a role in ensuring this and, if so, what should that be?

The Government can play a role in facilitating a broad consultation among stakeholders, but it should not attempt to draft a standard of its own. Standards need to work for a variety of use cases and at a global scale. If the Government drafted a set of UK standards, it would risk the standard either omitting issues important in other jurisdictions or not being adopted at all if compliance costs are unreasonably high for the UK market. That would discourage model training and development in the UK. We do think, however, the Government could serve a convening role on global standards more broadly.

Question 16. Does current practice relating to the licensing of copyright works for AI training meet the needs of creators and performers?

- No answer

Question 17. Where possible, please indicate the revenue/cost that you or your organisation receives/pays per year for this licensing under current practice.

- No answer

Question 18. Should measures be introduced to support good licensing practice?

- No answer

Question 19. Should the government have a role in encouraging collective licensing and/or data aggregation services?

- No

Question 20. If so, what role should it play?

We do not think that collective licensing should be encouraged nor is it appropriate. Mandating licenses upfront would effectively shift the system to an opt-in model, where AI developers must secure permissions before training models. This would slow AI development in the UK, limit access to training

data, and introduce complex negotiations, ultimately restricting innovation and the broader benefits AI will generate, undermining the Government's objectives.

Question 21. Are you aware of any individuals or bodies with specific licensing needs that should be taken into account?

- No answer

Question 22. Do you agree that AI developers should disclose the sources of their training material?

While transparency can be helpful for fostering trust in AI systems, it should be achieved in a way that is practical, proportionate, and preserves legitimate competitive interests and trade secrets. As described previously, a system that requires greater transparency for AI developers would ultimately intensify rights holders' demands for mass data licensing, with a range of adverse impacts on the UK's AI ecosystem.

Many AI developers, including OpenAI, already provide insights into how models are trained – often through documents like "model cards" or "system cards", which describe the categories of data used, how the model was evaluated, and the measures taken to protect privacy, mitigate risks and ensure the model operates safely. These disclosures offer transparency into AI development to support developers and researchers in better understanding AI models, without exposing proprietary methods. Any transparency mandates should build on existing practices like these, rather than impose unnecessary and redundant obligations.

Importantly, transparency requirements for AI training data should remain at a high level, allowing AI developers to explain in narrative form their sources of and their general approach to data collection. Requiring more detailed reporting, such as individual works or web URLs, would raise significant concerns related to model developers' trade secrets. AI developers invest heavily in refining their data and training processes, and publicising this knowledge would alter competitive dynamics between AI developers and could have national security implications (see https://www.belfercenter.org/publication/artificial-intelligence-and-national-security). A more balanced approach would involve aggregate, non-specific disclosures that maintain accountability while avoiding unnecessary competitive harms or administrative burdens.

OpenAI notes that under the EU AI Act, AI providers will have to provide a level of transparency into the data that was used to train AI models. While discussions are still ongoing at the time of consultation close, the UK should not introduce transparency requirements that would go beyond EU requirements, as that would undermine the UK's competitive position vis-a-vis the EU.

Question 23. If so, what level of granularity is sufficient and necessary for AI firms when providing transparency over the inputs to generative models?

See answer to Question 22.

Question 24. What transparency should be required in relation to web crawlers?

- No answer

Question 25. What is a proportionate approach to ensuring appropriate transparency?

See answer to Question 22.

Question 26. Where possible, please indicate what you anticipate the costs of introducing transparency measures on AI developers would be.

See answer to Question 22.

Question 27. How can compliance with transparency requirements be encouraged, and does this require regulatory underpinning?

If the Government imposed a regulatory requirement for transparency, the requirement should align with international standards, such as the EU AI Act's transparency rules, and industry norms such as model cards. If the UK requires AI developers to disclose more sensitive information than other jurisdictions, developers are likely to deprioritize the market.

Question 28. What are your views on the EU's approach to transparency?

We are actively engaged with the EU AI Office about the substance of the AI Act's transparency obligations and we have recommended that the AI Office crafts an approach that is not based on overly burdensome disclosures, as discussed in our response to Question 22. For example, we have suggested allowing providers to disclose approximate data sizes (e.g., ranges or rounded numbers) rather than precise figures, in order to protect competitive information and ensure clarity for the public. We have also advised caution when requiring disclosure of information — like demographic or regional characteristics — that may not be readily identifiable within the training data absent complex and expensive analyses. We have also raised concerns with detailing per-modality data sizes for different acquisition methods, suggesting instead that AI providers prepare narrative explanations of their data acquisition methods and contents, in line with AI Act Recital 107.

Question 29. What steps can the government take to encourage AI developers to train their models in the UK and in accordance with UK law to ensure that the rights of right holders are respected?

The AI Opportunities Action Plan includes strong proposals for strengthening the country's AI ecosystem, including ways of building more AI infrastructure, expanding access to training data, and accelerating AI adoption in both the public and private sectors. Each of these components is important for encouraging the growth of the UK's AI ecosystem, and for attracting international investment in the UK. As the Prime Minister has said, the Government won't be able to reach its economic growth targets without more fully embracing AI. Focusing on implementing these policies will be key for encouraging UK AI development,

including model training.

However, AI development is inherently international, with companies making strategic decisions on where to train models based on regulatory clarity, infrastructure cost, and ease of compliance. To attract AI developers, the UK's legal framework, its planning approach to datacenters, and its energy costs must be competitive with other leading jurisdictions. Building datacenters is already a costly investment, and complex, expensive compliance requirements will give AI developers incentives to train models elsewhere. Looking in isolation at copyright, a well-designed system that minimises administrative complexity and provides clear legal pathways for AI training will give firms confidence to operate in the UK rather than seeking simpler alternatives abroad.

Question 30. To what extent does the copyright status of AI models trained outside the UK require clarification to ensure fairness for AI developers and right holders?

Copyright is inherently territorial, and each country has the right to craft exceptions and limitations to copyright law, including for AI development, subject to international agreements like the Berne Convention. Attempting to "clarify" UK copyright law in a way that restricts lawful activities in other countries would not only violate fundamental principles of copyright law, but also infringe on national sovereignty.

This would also be inconsistent with the Government's goals to become a leader in AI, and would mean asking AI startups to navigate unprecedented questions of international law before they train their first model. Consider a developer training an AI model in Japan, where a legally accessible work has different rightsholders in Japan and the UK. Should they pay the Japanese rightsholder for reproduction rights that are not infringed under Japanese law, or the UK rightsholder for rights that do not apply in Japan? If the same developer later engages in training operations in the United States and uses the same work there, would they need a separate license? Rather than introducing unprecedented uncertainty, the UK should focus on providing a stable copyright framework that fosters AI innovation and respects international law.

Question 31. Does the temporary copies exception require clarification in relation to AI training?

No. Changes to this well-established framework are unnecessary, would create legal uncertainty, and may have unforeseen consequences for other industries.

Question 32. If so, how could this be done in a way that does not undermine the intended purpose of this exception?

- No answer

Question 33. Does the existing data mining exception for non-commercial research remain fit for purpose?

- No answer

Question 34. Should copyright rules relating to AI consider factors such as the purpose of an AI model, or the size of an AI firm?

- No answer

Question 35. Are you in favour of maintaining current protection for computer-generated works? If yes, please explain whether and how you currently rely on this provision.

- No answer

Question 36. Do you have views on how the provision should be interpreted?

- No answer

Question 37. Would computer-generated works legislation benefit from greater legal clarity, for example to clarify the originality requirement? If so, how should it be clarified?

- No answer

Question 38. Should other changes be made to the scope of computer-generated protection?

- No answer

Question 39. Would reforming the computer-generated works provision have an impact on you or your organisation? If so, how? Please provide quantitative information where possible.

- No answer

Question 40. Are you in favour of removing copyright protection for computer-generated works without a human author?

- No answer

Question 41. What would be the economic impact of doing this? Please provide quantitative information where possible.

- No answer

Question 42. Would the removal of the current CGW provision affect you or your organisation? Please provide quantitative information where possible.

Removal of protection for CGWs would not directly affect OpenAI. OpenAI does not claim copyright over generated outputs, and our Terms of Service for users in the UK and elsewhere state that “As between

you and OpenAI, and to the extent permitted by applicable law, you (a) retain your ownership rights in Input and (b) own the Output. We hereby assign to you all our right, title, and interest, if any, in and to Output."

See https://openai.com/policies/eu-terms-of-use/.

Question 43. Does the current approach to liability in AI-generated outputs allow effective enforcement of copyright?

- No answer

Question 44. What steps should AI providers take to avoid copyright infringing outputs?

As we explained in our response to Question 5, the Government should focus its attention on limiting the risks of copyright infringing outputs from AI models. Infringing outputs present clear, identifiable economic harm to creatives, and addressing them will not present risks to the UK's AI ecosystem like placing limitations on text and data mining. One approach would be to create a broad copyright exception for text and data mining that requires AI developers take reasonable measures to prevent infringing outputs.

AI developers have a variety of tools that can help prevent infringing outputs. For example, developers can take steps when collecting and curating training data to remove duplicate and near-duplicate copies of works, which reduces the likelihood of a model memorizing and potentially generating outputs that resemble that work. Additionally, during the post-training process, AI developers can teach a model to identify and refuse user requests to recite or generate copyrighted material. Model developers can also build classifiers to identify and block undesirable outputs as they are generated in real-time.

While these measures significantly reduce the risk of verbatim reproductions, they must also be calibrated to not over-block lawful uses. Not every reproduction is infringing — many legitimate outputs will make use of copyrighted materials pursuant to the UK's copyright exceptions, such as when a model quotes a sentence from a book to teach a user about a writing technique. Rigid approaches like strict output filters will restrict lawful responses and may not scale for models trained on trillions of points of data.

Responsible AI developers have a strong incentive to prevent infringing outputs because their models are designed to create new material. However, sometimes users take deliberate steps to evade model restrictions and intentionally generate infringing outputs. AI companies can and should take steps to anticipate and prevent this behavior, but determined users will be able to circumvent them in some cases, much as users can evade e-book DRM by taking screenshots of the open book file. In such cases, liability should fall on the infringing user—not the developer who acts in good faith to prevent such outputs.

Question 45. Do you agree that generative AI outputs should be labelled as AI generated? If so, what is a proportionate approach, and is regulation required?

OpenAI recognises the growing concern around the transparency of AI-generated content and agrees

that provenance solutions can play a key role. It is important, however, to craft solutions that are both effective and sufficiently broad to adapt to evolving AI technologies.

At this time, provenance solutions do not work for text. 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. As we described in a blog post about provenance (https://openai.com/index/understanding-the-source-of-what-we-see-and-hear-online/), we are not aware of labeling methods for text that are both accurate and cannot be circumvented by bad actors, such as by rewording text using another generative model.

For other modalities, we believe the most workable and sustainable approach is a simple, baseline requirement that clearly indicates whether AI was used to generate content and which AI developer's technology was involved. More detailed disclosures – such as the number of prompts used or the percentage of human involvement – are unlikely to remain practical as AI systems become more sophisticated and as user interactions with AI grow more complex.

Question 46. How can government support development of emerging tools and standards, reflecting the technical challenges associated with labelling tools?

The Government should avoid development of UK-specific standards on labeling. Article 50 of the EU AI Act requires AI providers to label AI-generated outputs by August 2026, and developers with a European presence will likely comply with whatever standards the AI Office determines are sufficient. The Government should encourage efforts to unify around a single set of globally recognised standards, rather than risking fragmentations across jurisdictions.

Separately there is already significant work within private industry on this issue, through bodies such as through the Coalition for Content Provenance and Authenticity (C2PA), of which OpenAI and other major AI developers are members. AI developers recognize the value of labeling standards, and this work is expected to continue independently, alongside any government efforts in this space.

Question 47. What are your views on the EU's approach to AI output labelling?

See our answers to questions 45 and 46.

Question 48. To what extent would the approach(es) outlined in the first part of this consultation, in relation to transparency and text and data mining, provide individuals with sufficient control over the use of their image and voice in AI outputs?

While we share concerns about the potential for bad actors to use generative AI systems to impersonate individuals, copyright law is the wrong legal framework to tackle this issue. Copyright protects the artistry in creative works, not physical attributes of the likeness or voice of performers in those works.

As this consultation notes, the UK already has legal protections that can address unauthorised digital replicas, including the tort of passing off, performers' rights, sound recording rights, and data protection laws. These frameworks already provide mechanisms to address unauthorised digital replicas, and targeted legislation could fill in any gaps. OpenAI has supported digital replica legislation, such as the United States's NO FAKES Act, and we would be happy to share our experiences with the Government in that regard.

Question 49. Could you share your experience or evidence of AI and digital replicas to date?

Like other applications of generative AI, responsible uses of digital replicas unlock new possibilities for communication, creativity, and efficiency. For example, we partnered with Spotify to power their new voice translation feature using our models, empowering participating podcasters to translate and dub English episodes into other languages (https://www.theverge.com/2023/9/25/23888009/spotify-podcast-translation-voice-replication-open-ai). Like other socially beneficial uses of digital replicas, our partnership allows podcasters to communicate directly to new audiences using a voice authentic to them.

While legitimate digital replicas have significant benefits, we have taken care as to potential risks. As we explained in the system card for our video-generating model, Sora, (https://openai.com/index/sora-system-card/), we built a stack of policies and tools to minimize the possibility Sora could be used to improperly impersonate individuals. For example, we monitor and block or rewrite prompts seeking to depict public figures, as well as run image output classifiers to detect whether generated video contains recognizable public figures (or human figures at all). While some depictions of humans are permitted by Sora, we currently monitor and block any generated videos that are based on uploaded images or videos of real minors, and we use a multi-tiered approach to detecting and limiting outputs of videos that are detected as sexually explicit or suggestive. Additionally, we limited the ability to generate content based on an uploaded image or video to a small group of users, subject to close monitoring, so that we could evaluate the effectiveness of our mitigation techniques and how creative professionals actually use image and video uploads in a video-generating model. Finally, we invested significant time into our work on provenance, so that if a user avoids our mitigations and generates a deepfake video, we are better able to identify that the video was generated using our services and how it was created. And we created a robust reporting, takedown, and enforcement flow that allows individuals to report misuse of their likeness and, if necessary, to take action against users who violate our usage policies.

While we have spent significant time protecting against misuse of Sora, safety measures can also impact societally beneficial uses. As we noted in our system card, a video depicting a protest can be used in a legitimate creative project, but it could also be presented as a real-world event as part of a disinformation campaign. In our early experience, users with the ability to generate videos of individuals through uploaded media have done so for benign purposes, and we have not seen significant misuse. Ultimately, there are difficult tradeoffs between stopping AI misuse and enabling free expression, and the Government will need to consider the appropriate balance in any legislative proposals.

In the context of synthetic voice, we have developed an AI model called Voice Engine that can generate realistic speech in a target voice using short sample audio of that person speaking. In designing safeguards, we have focused on measures to ensure the speaker of sample audio has given explicit and informed consent to replication of their voice. One method requires the input voice to read an AI-generated passage to confirm their active awareness and consent to provide input audio for the model. And as with video generated by Sora, we have invested in provenance solutions to ensure that if bad actors are able to avoid our safeguards and generate misleading audio, we are able to track the source of that audio.

Question 50. Is the legal framework that applies to AI products that interact with copyright works at the point of inference clear? If it is not, what could the government do to make it clearer?

Many inference uses are explicitly covered by exceptions for temporary copying, private use, or private study. The consultation suggests that under Option 3, other current copyright exceptions would not apply to any AI uses of copyrighted works. The Government should provide assurance that current copyright exceptions protect inference and that Option 3 will not allow rightsholders to reserve rights against it. If AI developers cannot rely on standard copyright exceptions to operate their models, they may exit the UK market.

Question 51. What are the implications of the use of synthetic data to train AI models and how could this develop over time, and how should the government respond?

Copyright rules that limit how developers can obtain and use human-created data will likely expand the use of synthetic data and affect the AI market. In some cases, this will advantage large firms, who will leverage their proprietary data to generate new forms of synthetic data not available to competitors.

Question 52. What other developments are driving emerging questions for the UK's copyright framework, and how should the government respond to them?

AI is rapidly evolving, with advancements in reasoning, data analysis, and agentic systems, which are unlocking new opportunities in research, productivity, and problem-solving. As these capabilities expand, it is essential that copyright law remains focused on its core purpose — protecting creative expression — while preserving enough flexibility to accommodate the broader role AI is playing in technological and societal progress.

For example, we've developed a new series of AI models designed to spend more time thinking before they respond. They can reason through complex tasks and solve harder problems than previous models in science, coding, and math. Through training, they learn to refine their thinking process, try different strategies, and recognize their mistakes. Enhanced reasoning capabilities will help people solve hard problems across a wide variety of domains, including science, coding, math, and similar fields. While we can't predict how quickly these models will scale, we expect to see more capable AI, in more places, sooner.

A restrictive copyright regime would harm model development and slow progress in areas that do not compete with or replace human creativity. For the UK to be a global leader in AI, it must adopt a forward-looking copyright framework rather than position copyright as a barrier to scientific advances resulting from large-scale data analysis. This is a prerequisite for UK leadership in AI innovation, and for capturing the economic benefits from AI investment and from the growth of the many sectors that leverage AI tools.