Copyright and AI: Response to Open Consultation

Table of Contents

Introduction

Thank you for the opportunity to contribute to this important consultation on the intersection of copyright and artificial intelligence (AI). My name is Kavana Ramaswamy, I am a PhD Graduand in Law at the University of Cambridge. As a legal academic, copyright law is deeply important both in my professional and personal capacity. My work often involves the use of copyright law for the licensing of my writing; research and writing form the bulk of my academic work. This will be directly impacted by any reform or change to copyright law. Additionally, I care about the rights of creative workers everywhere, whether this is done by writing research or literature, painting, music, or any other format.

The Open Consultation call states that the UK Government aims to (1) support right holders' control over content and retain the ability to be remunerated for its use, (2) support the development of world leading AI models in the UK, and (3) promote greater trust and transparency between the sectors. AI models are new technologies that alter the legal landscape substantially. Like any new technology, the training and use of AI models require new regulation to be harnessed and utilised in ways that provide a positive benefit to both the UK and to humanity as a whole. While the development of AI can be a force of good in the world, AI is a disruptive technology: it substantially alters what can be done in the world and exists currently in a legal vacuum. It is of utmost importance that such innovation be regulated to ensure that the disruption is creative, rather than destructive.

As such, the government has suggested certain courses of action in the open consultation on how copyright law ought to be reformed. I strongly believe that the suggested reforms will not adequately support the government's first aim: supporting right holders' control over content and retention of the ability to be remunerated for its use. The suggested reforms will effectively negate

the first aim in favour of the second, as opposed to striking a sustainable balance between the two. The suggested changes will also negate the third aim and instead, erode trust and transparency between the sectors. It will also have the effect of eroding many people's trust in the government, which I do not believe that the government intends as an outcome.

While I believe that the suggested reforms will not serve the aims of the government, I do believe that reform is necessary for addressing the gaps in the law and bring AI innovation within the regulatory framework. Furthermore, the suggested policy changes fail to account for problems in various licensing models, which will complicate matters further, should a general exemption for data mining be granted as the government intends. To this end, I have suggestions on what might be a better way to meet the aims of the government while also taking into consideration some other issues that are also at play in this context. I sincerely urge the government to take this response (and the responses of other creators directly impacted by this policy) seriously, as consequences of not doing so are significantly more dire than most people anticipate.

Ethics of Generative AI

While much has been said on ethical issues regarding AI more generally,the ethics of generative AI raises specific problems that need to be addressed in any policy that aims at regulating the sector.

How Generative AI Works

Generative AI does not "create" new work/art in the traditional sense. Rather, it provides an output by processing input (the data mined) and providing an output that uses the input data as synthesis material. Fundamentally, it operates by registering patterns in the input data and generating an output based on probability.[1] The output generated is therefore always a derivative of the input data: the most likely next word/pixel/sound from the given input and the mined data. The so-called "creation" will always be a probabilistic response to the input. As such, it would be a mistake to relate this to human creativity, which, at its best, is not derivative. It is important to distinguish between what we instinctively believe that AI is doing, viz. "creating" works of art or "understanding" human language, and what the generative AI is actually doing, i.e., processing and regurgitating mined data.[2] However, even in the event that AI development progresses to the point where AI models operate differently, the issues raised in this response (such as issues of commercialisation and use) would still be relevant.

The derivative nature of generative AI results in several problems arising out of the output generated by them. Scholars have pointed out various problems in the use of generative AI for

1 Vassilka D Kirova and others, 'The Ethics of Artificial Intelligence in the Era of Generative AI' (2023) 21 Journal of Systemics, Cybernetics and Informatics 42, 44; Emily M Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021) 611 <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 24 February 2025.

2 Bender and others (n 1) 610–611.

everyday tasks; these include bias[3], misinformation[4], security risks, deep fakes, resource drains, and many others. Many of these risks arise from the nature of data mined as source material for generative AI to produce content; if a large portion of the source data contains misinformation/biased data, the AI will use probabilistic predictions to accurately mimic the misinformation/bias in the source.[5] Using AI generated data as additional source material for further training of AI models results in feedback loops where this bias is entrenched into the system.[6]

Problems of Expanding Generative AI

While some of these issues may be addressed by controlling for the quality and kind of data that the AI is trained on, there are other problems associated with increasing the quantities of data mined and processed for generative AI. This is not an unbidden good; like any new technology, we need to seriously consider the dangers involved in new technology before choosing to endorse, support, or subsidise the development and expansion of it. Certainly, we would not agree to expand the development of, say, biological weapons simply because they may afford new avenues for technological innovation and scientific research. Generative AI is no exception to this: we need to consider the risks of expanding access to source material for data mining by AI companies before implementing any changes to copyright law for this purpose.

Research suggests that increasing the size of AI models by expanding the datasets that they are trained increases the difficulty in implementing and enforcing transparency on the training data.[7] The government has stated that it aims to promote trust and transparency between the creative and AI sectors. This is a commendable goal, but it must be noted that the creative sector does not have anything to hide in this context. Promoting trust and transparency between the sectors assumes that the creative sector is being opaque in its interactions with the AI sector, which is simply not true. The existence and use of copyright law by the creative sector has nothing to do with transparency, but with ownership and control of creative work. The issue of transparency primarily arises in the context of AI companies using copyrighted works as training data for generative AI. This is a unidirectional issue of transparency, where the creative sector is largely unaware of what data is being used to train AI. For copyrighted work where no right has been granted for derivatives to be created from the work, AI companies using such data to train generative AI are clearly in violation of the copyright on such work; they further marginalise creative workers by failing to explicitly seek permission to use copyrighted work for the training of generative AI or remunerate creators for such use. Further, reports have indicated that companies may be accessing such data illegally[8]—

3 Timnit Gebru, 'Oxford Handbook on AI Ethics Book Chapter on Race and Gender' (arXiv, 8 August 2019) <http://arxiv.org/abs/1908.06165> accessed 22 February 2025; Jackie Kay, Atoosa Kasirzadeh and Shakir Mohamed, 'Epistemic Injustice in Generative Ai', *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2024) 686 <https://ojs.aaai.org/index.php/AIES/article/view/31671> accessed 24 February 2025.

4 Kirova and others (n 1) 46; Plamena Zlateva and others, 'A Conceptual Framework for Solving Ethical Issues in Generative Artificial Intelligence', *Electronics, Communications and Networks* (2024) 112–113 <10.3233/FAIA231182>.

5 Gebru (n 3) 2–24.

6 ibid 7–9.

7 Bender and others (n 1) 610.

8 Ashley Belanger, '"Torrenting from a Corporate Laptop Doesn't Feel Right": Meta Emails Unsealed' *Ars Technica* (6 February 2025) <https://arstechnica.com/tech-policy/2025/02/meta-torrented-over-81-7tb-of-pirated-books-to-train-ai-authors-say/> accessed 24 February 2025; Kate Knibbs, 'Meta Secretly Trained Its AI on a Notorious Piracy Database, Newly Unredacted Court Docs Reveal' *Wired* <https://www.wired.com/story/new-documents-

using databases such as shadow libraries—to circumvent copyright restrictions altogether. While transparency is important, it is crucial to understand that the obligation of transparency is entirely on AI companies, not the creative sector.

Additionally, training AI on large data sets involves substantial energy and environmental costs; the training and deployment of AI models involves the use of significant resources—such as electricity and water—and emits vast quantities of carbon dioxide.[9] While it may be argued that electricity may be provided for using renewable sources, it remains pertinent to question the need for diverting energy resources to expand data mining at a time when the UK (and the world at large) is experiencing a severe energy crisis. Over 6 million people in the UK are living in fuel poverty in 2025.[10] We must be wary of exponentially increasing demands on our energy networks for the sake of expanding AI tools. Such increased demands will inevitably raise the costs of consumption of energy, further entrenching inequality and energy poverty in the UK.

Copyright Issues

The government primarily suggests implementing a data mining exception allowing AI developers to train AI on any material they legally have access to, with right holders being allowed to reserve their rights, underpinned by supporting measures on transparency. This suggests an "opt out" system, where copyrighted works are, by default, available to AI companies for data mining. Creators will be required to explicitly opt out of having their material mined by AI companies. The burden of navigating this system would effectively fall on every individual creator to explicitly mark every single work produced as being restricted from use by AI companies for mining. The suggested policy blatantly ignores the power differences existing between large companies and individual creators, and further overwhelmingly sides with the large corporation that has both the means and the resources to effectively access data in an alternative "opt in" system and throws individual creators to the wolves.

For the elucidation of the problems with this approach, I will use examples of creative work licenced under "creative commons" licenses. Creative commons licenses are specific forms of copyright reservation where the right to access the work (i.e. the right to view/read/hear the work) is granted to the general public, but other rights (such as the right to edit or make derivative work, or the right to sell the work for remuneration) may be reserved by the creator of the work. As such, anything that affects work licensed under a creative commons license would also apply to the work had it been copyrighted and not licenced. I therefore use the least restrictive form of copyright license (the creative commons license) to demonstrate problems with the suggested policy that would automatically apply to stronger forms such as non-licenced copyrighted work. It is also pertinent to note that many creators and academics who licence work under creative commons licenses do so with a very specific intention on how the work ought to be used, and generally reserve rights to prevent the work from being used in unintended ways. Our rights as creators ought

unredacted-meta-copyright-ai-lawsuit/> accessed 24 February 2025.

9 Bender and others (n 1) 612–613; Pengfei Li and others, 'Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models' (arXiv, 15 January 2025) 1–6 <http://arxiv.org/abs/2304.03271> accessed 24 February 2025.

10 'Energy Crisis Timeline: How the Energy Crisis Unfolded' (*National Energy Action (NEA)*) <https://www.nea.org.uk/energy-crisis/energy-crisis-timeline/> accessed 24 February 2025.

to be respected, and changes to copyright law that override these rights will inevitably be very unpopular policies. Many noted artists have already registered their mistrust of the proposed policy.[11] Implementing the policy suggested by the government will have the impact of alienating millions of creators in the UK (academics, writers, painters, musicians, and many other artists) in favour of enhancing the profits of multi-national mega-corporations. I would sincerely urge the government to rethink this policy suggestion and listen to the needs of the individual people seeking to express and share their creativity with the country and the world over multi-national corporations that seek to steal the works of artists for profit.

Creative Commons Licenses

Creative Commons was founded as an alternative framework to copyright.[12] At the heart of the creative commons (CC) community is the desire of creators and artists around the globe to share our work for the benefit and growth of humanity. CC licenses operate within the framework of copyright law, and allow for the free sharing of creative work with people around the world, while reserving certain rights to the creator. This empowers creators to choose what rights they wish to reserve, ensure that the work they share is used by people across the globe in ways that they intended, and ensure that they are credited with creating the work. In academia, the open access movement seeks greater accessibility to education and learning by allowing free and open access to academic information and publications, eliminating financial/legal/technical barriers to accessing information.[13]

CC licenses specifically allow creators to reserve specific rights to work: attribution, commercialisation, and editing. These are usually denoted by the letters 'BY' for attribution, non-commercial ('NC') for the right to commercialisation, share-alike ('SA') to denote that sharing of the work must be done under similar CC licenses, and no-derivatives ('ND') to indicate that the author does not authorise any distribution of derivative work made from editing their work. These rights are substantially differentiated from work submitted to the public domain, usually indicated by 'CC0' where the creator dedicates the work to the public domain. CC licenses additionally operate at three levels: licence text that is human-readable, lawyer-readable, and machine-readable.[14] The open access culture fostered by the CC movement should not be mistaken as permissiveness for the use of artwork for profit; rather, it should be understood as a human-centred approach to empower creators and the creative community over and above the commodification of art. In the subsequent paragraphs, I elucidate some of the problems raised by the government's suggested policy in the context of CC licences.

11 Vanessa Thorpe, Vanessa Thorpe Arts and media correspondent, 'Don't Gift Our Work to AI Billionaires: Mark Haddon, Michael Rosen and Other Creatives Urge Government' *The Observer* (23 February 2025) <https://www.theguardian.com/technology/2025/feb/23/dont-gift-our-work-to-ai-billionaires-mark-haddon-michal-rosen-and-other-creatives-urge-government> accessed 23 February 2025.

12 Nick Scharfi, 'Creative Commons-Ense? An Analysis of Tensions between Copyright Law and Creative Commons' 12 Journal of Intellectual Property Law & Practice 376.

13 'What Is Open Access? | Open Access' <https://www.openaccess.nl/en/what-is-open-access> accessed 24 February 2025; 'Berlin Declaration' <https://openaccess.mpg.de/Berlin-Declaration> accessed 24 February 2025.

14 Lawrence Lessig, 'The Creative Commons Commentary' (2004) 65 Montana Law Review 1, 11.

ND/SA Clauses

The proposed "opt-out" system interacts in questionable ways with work shared under a CC licence, effectively dismantling the rights of authors under this system. Fundamentally, as a system that processes input for the explicit purpose of recognising patterns and generating derivative work, generative AI would *always* violate the ND clauses on CC contracts. Even if the AI is never used to make something that looks similar to or substantially like the original work, the processing of the CC work by the AI is, by itself, use in violation of the ND clause as the intended purpose of the data mining (rather than a person merely accessing the work) is to produce derivative work through predictive pattern-matching. Further, it is unclear that mined data can ever be reversed. This places creative artists who wish to build a community on shared human values at a substantial disadvantage to the companies seeking to profit from the unlicensed theft of their work.

The SA clause will also be violated by any AI company that does not use the same or a more permissive CC license for any output generated by the models. While there are AI companies that are non-profits, they may still require remuneration for the generation of content, or have more restrictive copyright clauses than the original if the law permits it.

NC Clauses

Under the proposed data mining exemption by the government, companies may use any data made available on the internet as source material for data mining for any purpose, including commercial purposes.[15] This essentially negates the 'NC' clause of CC licences, ensuring that large AI companies have full access to CC content without having to seek specific licences to use these for commercial purposes despite the explicit prohibition.

This is further complicated by the existence of AI companies that are not-for profit, because not only do they sometimes change their structure and become for-profit companies,[16] but the data they mine and their models may be used by other AI companies[17] (who may be operating for profit) and thus effectively violate the NC clauses on CC work. Large AI companies have been seen to act hypocritically, claiming that the theft of their own intellectual property is "inappropriate" while having trained their own AI models on data taken from creators without their consent.[18]

15 Secretary of State for Science, Innovation and Technology, 'Copyright and Artificial Intelligence' (Intellectual Property Office) Consultation E03259461 12/24 para C.1 <https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence/copyright-and-artificial-intelligence> accessed 21 February 2025.

16 For example, OpenAI, which began as a not-for-profit company, is considering becoming a commercial for-profit enterprise (Dan Milmo and Dan Milmo Global technology editor, 'Why Is OpenAI Planning to Become a For-Profit Business and Does It Matter?' The Observer [26 September 2024] <https://www.theguardian.com/technology/2024/sep/26/why-is-openai-planning-to-become-a-for-profit-business-and-does-it-matter> accessed 24 February 2025).

17 For example, OpenAI has claimed that DeepSeek used its models to train the DeepSeek AI (Jess Weatherbed, 'OpenAI Has Evidence That Its Models Helped Train China's DeepSeek' [2025] *The Verge* <https://www.theverge.com/news/601195/openai-evidence-deepseek-distillation-ai-data> accessed 24 February 2025).

18 'OpenAI Says DeepSeek May Have "inapproriately" Used Its Data' *NBC News* (30 January 2025) <https://www.nbcnews.com/tech/tech-news/openai-says-deepseek-may-inapproriately-used-data-rcna189872> accessed 24 February 2025; Eleanor Olcott and Cristina Criddle, 'OpenAI Says It Has Evidence China's DeepSeek Used Its Model to Train Competitor' *Financial Times* (29 January 2025) <https://www.ft.com/content/a0dfedd1-5255-4fa9-8ccc-1fe01de87ea6> accessed 24 February 2025.

BY Clauses

Any derivative work (or indeed, an exact copy of the original without edits) that does not credit the original author/creator of the work shared under a CC-BY licence is in violation of the terms of use of such work. Should an AI generate an image that is similar to or a copy of an existing work of art without crediting the author, it will be in violation of the original licence. Many image generation websites replicate copyrighted works of art without any statement/attribution to the original author. These may be used by unsuspecting users for whatever purposes the AI company licences it, again, without credit to the creator. Many creators depend on attribution for publicity; even if the art is made available to the public to access/view without cost, most artists nonetheless stamp the work of art with their signature, or demonstrate their authorship of the work in some way.

AI that generates text also regularly generates text without citing sources; often one is required to specifically request sources for the text cited for the authorship to be traced. Even then, given the probabilistic nature of AI generation, the attribution is often mistaken or complete nonsense generated based on patterns rather than the actual origin of the source data. Failure to ensure proper attribution (as is the case of most AI models in use currently) harms creators by denying them credit for their contributions, even when such contributions are used by the AI companies and their users. When humans deliberately fail to give appropriate credit to sources material used in creative work, we understand this to be plagiarism, and a violation of academic/creative integrity. We ought not hold AI companies and the technologies that they unleash onto the world to a reduced standard of ethics, particularly when a large number of mega-corporations that have the funds, resources and means to prevent such violations are now involved in the creation and deployment of these models. We most certainly should not allow tech companies a free licence to violate the rights of individual creators who may or may not have the funds to fight large corporations on the violation of our rights.

Additional Issues

The problem of a general data mining exception granted to AI companies is that every work of art licenced under a CC license on the internet today – which includes millions of works of research and art, from research papers, fictional stories, to images and paintings, photographs, and music, becomes retro-actively available to AI companies to use in whatever way they wish, in blatant violation of the CC licences used by the creators. Requiring individual creators to find every single piece of artwork they have ever created or shared on the internet in order to restrict access to AI companies places an onerous burden on individual creators, many of whom do not have the resources, means, knowledge or time to adequately ‘opt out’ of each of these works being mined by AI companies. Instead, AI companies—many of whom are large multi-national mega-corporations—have both the resources, means, time (in the form of thousands of employees) and ability to seek permissions for all the copyrighted and CC-licenced works that they wish to use in their data sets, and additionally remunerate creators for the use of these in generative AI training. While AI companies may not currently have dedicated employees to do this work, requiring them to do so would additionally create jobs for people in the UK, supporting the government’s goals of greater growth and expansion of AI development within the UK.

All of the problems highlighted above apply equally to copyrighted works of creators, as full copyright is the stronger form of intellectual property protection.

Thus, we see that the suggested changes to copyright law to allow for a general data mining exemption will not meet the aims of the government in as much as the government aims to support right holder's control over our content and our ability to be remunerated for its use. While it may support the development of AI models, it is certain to erode any trust in AI companies, and very likely to erode the trust of the entire creative sector in the current government.

Responses to Questions

To actually support creators in the UK and seek greater transparency and trust between the AI and creative sectors, I suggest instead the following, in response to the some of the questions posed by the government:

Question 1. Do you agree that option 3 is most likely to meet the objectives set out above?

No. As explained in the preceding paragraphs of this submission, option 3 erodes the rights available to creators under various copyright and CC licences, and is most likely to strip, rather than support, right holders' control of their content and their ability to be remunerated for their work and its use. It will also inevitably erode both the trust of creators in AI companies and the trust of the entire creative sector in the current government. The only objective it serves is supporting the development of AI, at the cost of placing significant burdens on creators.

Question 2. Which option do you prefer and why?

I prefer Option 1: Strengthen copyright requiring licensing in all cases. This is because the power imbalance that exists between individual creators and large AI companies requires the government to seek to support the individual right holder against the coercive monetary power of the AI companies in order to attain a power balance and an actual building of trust between the two. AI companies have already been seen to pay lip service to intellectual property rights; granting them general exceptions to copyright as a whole will only embolden them further and result in individuals having no option to be remunerated for work. At any rate, the proposed reform must ensure that all forms of copyright, including CC licences, are required to be respected by AI companies.

Question 3. Do you support the introduction of an exception along the lines outlined above?

No, as stated earlier in the submission and in question 1, option 3 erodes the rights available to creators under various copyright and CC licences, and is most likely to strip, rather than support, right holders' control of their content and their ability to be remunerated for their work and its use. It will also inevitably erode both the trust of creators in AI companies and the trust of the entire creative sector in the current government. The only objective it serves is supporting the development of AI, at the cost of placing significant burdens on creators.

Question 4. If not, what other approach do you propose and how would that achieve the intended balance of objectives?

I propose option 1 as the best way forward that would achieve the intended balance of objectives. It is important to recognise that balance requires an understanding the balance of power and that true balance requires power *imbalances* to be corrected. Currently, there is an imbalance of power with large AI companies having the means to override individual rights entirely. Balance requires strengthening the rights of individuals against this, and ensuring that our rights are respected by AI companies.

Question 5. 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.

As an academic, an exemption along these lines would have severe negative impacts on my work. Specifically, because AI models do not provide appropriate citations and credit to the sources of text generated, many academics risk being plagiarised and losing out on impact metrics that are very important to our work and career advancement. Additionally, it is likely to harm the research community by disincentivising the sharing of research work and the consequent advancement of knowledge that this fosters. Where academics are currently moving towards open access, fear of our work being stolen and plagiarised by AI companies and their users may discourage people from publishing research in open access formats, stymieing research networks instead of fostering them.

Question 6. What action should a developer take when a reservation has been applied to a copy of a work?

I do not support a reservation system, but a licence system where works are reserved by default and only licenced when explicitly granted. Developers should seek to obtain a licence to use the work directly from the creator/publisher of such work. Many publishers will be happy to licence the work to AI, understanding the benefits of a shared space of innovation and technology. However, where the creator does not wish to licence a work, this needs to be respected by the developer. AI does not require every single piece of work to be mined in order to be effective at generating outputs that are useful to the user.

Question 7. What should be the legal consequences if a reservation is ignored?

A reformed law should include penalties, including exemplary compensation to the creators for violation of copyright. This requires exemplary damages because regular damages will incentivise large companies (many of which are worth billions of dollars) to regularly violate copyright because the cost of paying the damages exceed the cost of obtaining a licence. This is not to say that seeking licences would be onerous to the company, but that the profit margins would be sufficient incentives to violate copyright without exemplary damages.

Question 8. 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. The cost of introducing machine readable formats would not be significant, given particularly that such systems already exist in the creative commons licensing system. Machine readability ensures that CC-licensed works can be specifically searched for on google, for instance, among

other benefits. Expanding this to allow for a machine-readable opt-in "AI licence" for those wishing to allow AI companies to use their material will not be difficult or expensive.

Question 9. Is there a need for greater standardisation of rights reservation protocols?

Yes. Standardisation of copyright already exists to a certain extent, and CC licences are used across the globe. The emergence of AI requires expansion of this standardisation to create an 'opt-in' systems for creators to licence work to AI companies, or request AI companies to contact them directly for licensing.

Question 10. How can compliance with standards be encouraged?

As mentioned in question 7, exemplary damages can encourage compliance.

Question 11. Should the government have a role in ensuring this and, if so, what should that be?

Yes, the government's role in ensuring this is to enact legal and regulatory frameworks that support the rights of individual creators in order to ensure a true balance of power.

Question 12. Does current practice relating to the licensing of copyright works for AI training meet the needs of creators and performers?

No. The current practice of licensing of copyright has not protected the needs of creators and performers because it has encouraged AI companies to use all and any data that they have access to as source data for AI training. This needs to be discouraged and the rights of creators and performers needs greater protection.

Question 14. Should measures be introduced to support good licensing practice?

Yes

Question 15. Should the government have a role in encouraging collective licensing and/or data aggregation services? If so, what role should it play?

Yes. The role of the government in each case is to support the individual creators in forming unions to have collective bargaining power against AI companies to negotiate better terms for licensing of works.

Question 16. Are you aware of any individuals or bodies with specific licensing needs that should be taken into account?

Yes. Individual creators who are licensing research/art under CC licences need our rights to be protected by law, particularly in the context of AI companies violating our rights under the CC licences. Academics need to be assured that our rights of attribution will be respected and that our research will not be used to train AI models without our explicit consent, in order to continue to support an environment of open learning and open access for research work.

Question 17. Do you agree that AI developers should disclose the sources of their training material?

Yes. This is essential for transparency, and to have accountability for potential harms by AI generated content (eg misinformation, bias, etc)

Question 19. What transparency should be required in relation to web crawlers?

Web crawlers need to be open for inspection of code, and the code needs to respect machine-readable licence terms/restrictions.

Question 20.What is a proportionate approach to ensuring appropriate transparency?

Ensure that the data used to train AI is open for inspection by independent bodies on request.

Question 21. Where possible, please indicate what you anticipate the costs of introducing transparency measures on AI developers would be.

The cost of introducing transparency measures on AI developers would not be tiny, but these costs are essential for the maintenance of transparency between individual creators and AI companies. Much like the right to privacy imposes costs on both large and small companies alike in order to be GDPR compliant, the cost of development of transparency measures must be understood to be a part of the business costs of creating AI in a sustainable manner

Question 22. How can compliance with transparency requirements be encouraged, and does this require regulatory underpinning?

Exemplary damages for violation of transparency requirements, encoded into the legal framework as minimum damages for violations.

Question 24. 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?

Have a regulatory framework which addresses the gaps in the law. Encourage the creation of a machine-readable code that creates the 'opt-in' system for individuals to control how the data is used. Provide incentives for companies that operate as not-for-profits and comply with the regulatory framework, and demonstrate transparency in their operations.

Question 25. 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?

Clarification to be issued that the use of these AI models in the country will still be subject to UK law.

Question 26. Does the temporary copies exception require clarification in relation to AI training?

Yes – temporary copies may still be used as AI training data in violation of the copyright of the work that is temporarily copied. Further, the use of a temporary copy as AI training data effectively becomes a 'permanent' data point in the AI so trained, and thus cannot be considered exempted under the temporary copy exemption.

Question 28. Does the existing data mining exception for non-commercial research remain fit for purpose?

No, specifically because there are AI companies that are currently not-for-profit but may convert to commercial enterprises in the future. The data mining exemption for non-commercial research cannot include AI models, because the model so trained may in future become commercial, unlike specific/individual research projects.

Question 29. Should copyright rules relating to AI consider factors such as the purpose of an AI model, or the size of an AI firm?

No. Like the General Data Protection Regulation (GDPR), copyright protection is required to protect individual rights against companies and service providers of all sizes and purposes.

Question 38. Does the current approach to liability in AI-generated outputs allow effective enforcement of copyright?

No. Enforcement must shift to exemplary damages for violations of copyright in order to ensure transparency and appropriate oversight on licensing and development of the AI by companies.

Question 39. What steps should AI providers take to avoid copyright infringing outputs?

AI companies need to include as a part of their business model dedicated teams for licensing and obtaining source data for training AI models.

Question 40. Do you agree that generative AI outputs should be labelled as AI generated? If so, what is a proportionate approach, and is regulation required?

Yes. Regulation is required to ensure that AI companies comply with this provision. It is appropriate for the purpose it serves, i.e. alerting a user to the source of the data and the possibility that the data may not be an accurate representation of reality, should the data be referring to something in the world.

Question 41. How can government support development of emerging tools and standards, reflecting the technical challenges associated with labelling tools?

The government can create a specific body for the development of digital tools and standards to provide, for e.g., machine readable opt-in licences for individuals to use.

Question 43. 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?

Only strengthening copyright rules will provide individuals with sufficient control over the use of their image and voice in AI outputs. Additionally, we need to ensure regulations to prevent privacy violations on this front, such as misrepresentation of an AI output as a recording of a real person's statements.

Question 45. 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?

No. The government needs to ensure that AI products that interact with copyright works have the licence to use those works for derivative AI generation, and additionally respect the rights of authors to be credited for output that is substantially similar to their original work.