



**BEFORE THE
UNITED KINGDOM INTELLECTUAL PROPERTY OFFICE**

**Consultation on Copyright and Artificial
Intelligence**

25 February 2025

The Copyright Alliance and its members appreciate the opportunity to submit these comments in response to the Consultation on Copyright and Artificial Intelligence (Consultation) launched by the United Kingdom Intellectual Property Office (U.K. IPO). We write to offer the views of the U.S. creative community and to encourage the U.K. Government to continue championing the rights of creators, copyright owners, and creative professionals by rejecting the proposal for a text-and-data (TDM) mining AI exception in U.K. copyright law.

As we explain in detail below, the proposed TDM AI exception would severely undermine copyright. It is problematic for a variety of reasons, including that it would dramatically reduce rights holders' rights and incentives to create and to widely disseminate their creations, violate the U.K.'s international intellectual property (IP) obligations, and run counter to other countries' approaches to AI, including the United States.

The Copyright Alliance is a non-profit, non-partisan public interest and educational organization representing the copyright interests of over 2 million individual creators across the globe and over 15,000 organizations in the United States, across the spectrum of copyright disciplines.¹ The

¹ A full list of our organizational members is available on our website at: <https://copyrightalliance.org/about/who-we-represent/>.

Copyright Alliance is dedicated to advocating policies that promote and preserve the value of copyright, and to protecting the rights of creators and innovators. The individual creators and organizations that we represent rely on copyright law to protect their creativity, efforts, and investments in the creation and distribution of new copyrighted works for the public to enjoy.

The Copyright Alliance and our members support responsible, ethical, and respectful development and use of AI technologies. An AI ecosystem that meets these criteria is one that (1) values and respects the rights of creators and copyright owners and the importance of the copyrighted works they create and (2) does not make those rights and works subservient to the interests of AI companies. Copyrighted works are a fundamental part of any AI development. The countless movies, music, books, articles, visual art, and other creative works that have made the U.K. a cultural, artistic, and information powerhouse are made possible because U.K. copyright laws have incentivized and fostered human creativity and labor. It is vital that the U.K. Government continue its steadfast support of its £124.8 billion creative sector and the copyright laws that support it.

Copyright law is also the foundation of the rich history of U.K.-U.S. trans-Atlantic creative collaborations and mutually supportive creative sectors, which have facilitated strong economic growth, job creation, a robust trade relationship, and invaluable artistic and cultural achievements. It is what makes both of our countries global cultural leaders. Despite all this, the U.K. Government seems willing to gamble away its cultural heritage and its long-held position as a recognized leader in the creative industries for the mere possibility of becoming a leader in AI technologies. With the global proliferation of AI technologies, the U.S. creative community and other stakeholders strongly believe that it is crucial, now more than ever, for the U.K. Government to continue championing strong intellectual property (IP) rights and to send the message to the world that strong and effective copyright protection promotes, rather than hinders, AI innovation, especially in the case of generative AI technologies (GAI).

In the Consultation, the U.K. Government states that it believes in a future where the creative industry is a part of AI innovation, indicating that it wants “to ensure that both the AI and creative industries can share in the benefits of AI and that both sectors are able to grow

together.” It acknowledges that copyright law enables the licensing of creative works for AI use. Despite these proclamations, the Consultation departs from that vision by proposing a TDM exception for AI uses (Proposed TDM Exception) that would lay to waste the rights of creators and copyright owners by effectively subjugating them to subsidize AI technologies. This would decimate the U.K.’s long and rich history of cultural and artistic leadership.

In the Consultation, the U.K. Government sets forth three objectives as its reasons for the Proposed TDM Exception: (1) rights holders expressed concerns of a lack of control over their works and the ability to license and receive remuneration for the use of their creative works by AI models; (2) rights holders expressed concerns over a lack of transparency which has made copyright enforcement difficult; and (3) AI companies expressed concerns that there is a lack of clarity over how to legally access copyrighted works to use for AI training purposes.

Unfortunately, the U.K. Government wrongly concludes that the best way to achieve these objectives would be to enact the Proposed TDM Exception combined with affirmative opt-out, a legal access requirement, and transparency obligations. To the extent that U.K. policymakers think that providing rights holders with an option to opt out, imposing a legal access requirement, and establishing training dataset transparency obligations balances things out, that calculus is woefully incorrect (as we will discuss in more detail). This approach isn’t balanced, just, or reasonable—it’s an approach that is destructive and detrimental to the U.K.’s creative community. It is based on the incorrect premise that copyright law is an obstacle to the success of the U.K. AI industries, when in fact copyright law is an integral part of ensuring the success of the U.K. AI industries.

As explained in further detail below, licensing between rights holders and AI companies is achievable. Since the rise of GAI technologies a few years ago there has been more licensing deals. As we’ve learned from prior technological innovations like the copying machine, the VCR, the internet, digital music services, and many other new technologies, copyright laws did not end or hinder the development of these technologies—nor will it do so now with AI. Copyright and AI can progress successfully together without the need for a government-created exception. Indeed, if U.K. policy acts as a disincentive to the continued creation of high-quality,

human-authored works, it will also undermine the success of the U.K. AI industries as the quality of AI tools deteriorates rapidly without high-quality, human-authored works to train on.

Because the Proposed TDM Exception is both unwarranted and misdirected, the Copyright Alliance urges the U.K. Government to withdraw it, like the exception proposed by the previous government was rejected.² We address below some of the questions posed in the Consultation.

Sections B.4 Policy options & C.1 Exception with rights reservation

1. Do you agree that option 3 - a data mining exception which allows right holders to reserve their rights, supported by transparency measures - is most likely to meet the objectives set out above?

3. Do you support the introduction of an exception along the lines outlined in section C of the consultation?

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.

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.

No. The Proposed TDM Exception fails to meet the objectives of the Consultation. There already exists a framework that provides the necessary balance; one that provides the requisite control for creators and rights holders and requisite access for AI developers and deployers. That framework is copyright law and the associated free market licensing system.

Copyright law enables creators and rights holders to supply AI companies with flexible and responsive solutions for AI and tailor licensing and business models for GAI development. The ability of creators and rights holders to create works and enforce rights in those works is crucial, because it incentivizes the further creation and proliferation of high-quality creative works which form the basis for GAI development. Without copyrighted works, GAI technologies cannot

² HOUSE OF COMMONS CULTURE, MEDIA, AND SPORT COMMITTEE, CONNECTED TECH: AI AND CREATIVE TECHNOLOGY: GOVERNMENT RESPONSE TO THE COMMITTEE'S ELEVENTH REPORT OF SESSION 2022-23, 4 (Third Report of Session 2023-24) (Jan. 9, 2024), <https://committees.parliament.uk/publications/42766/documents/212749/default/>.

generate high-quality output. AI companies are also seemingly accepting this reality, as there are more licensing and partnership deals with rights holders being reached with each passing day.

Increasing numbers of copyright owners, particularly news, magazine, and academic publishers and image/media licensors are licensing their copyrighted works to AI companies for commercial AI uses and have been doing so for many years. *This shows that the market is working and there does not need to be any change in the law.* In the U.S., just a few public examples of recent licensing solutions, initiatives, partnerships, and agreements for AI use of copyrighted works include those launched from or created by [Authors Guild](#), [Created by Humans](#), [Dataset Providers Alliance](#), [Copyright Clearance Center](#), [Elsevier](#), [Getty Images](#), [Shutterstock](#), [Jstor](#), [Sage Journals](#), [Rightsify](#), [Universal Music Group](#), and other major media publishers including the [Associated Press](#),³ [Axios](#), [Condé Nast](#), [News Corp](#), [The Atlantic](#), [Vox Media](#), [Dotdash Meredith](#), [Fortune](#), [Time](#), [Entrepreneur](#), [The Texas Tribune](#), and [WordPress.com](#).

The AI-copyright licensing market has grown over time, with the number of GAI licensing and partnership deals continuing to further expand that market. But this growth will be stunted if the U.K. adopts the Proposed TDM Exception. Nobody disputes that AI companies and developers must pay for and invest in computer chips and cloud infrastructure. It is part of the cost of doing business. Copyrighted works are just as important, if not more important, than the chips and infrastructure. AI developers should be required to pay for the copyrighted works that they use to train their AI systems. The mere speculative future of a technology cannot be the only rationale to justify a technology-specific exception in light of the fact that strong copyright laws can and already have been shown to foster AI innovation. The fact that under current copyright laws AI licensing has been growing and will continue to do so, shows that the present system is working and that a government intervention that effectively requires rights holders to subsidize AI development would be a cataclysmic mistake. It is a mistake that would threaten not only the

³ See Matt O'Brien, *ChatGPT-maker OpenAI Signs Deal with AP to License News Stories*, ASSOCIATED PRESS, (July 13, 2023), <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a>; Matt O'Brien, *Google Signs Deal with AP to Deliver Up-to-date News Through Its Gemini AI Chatbot*, ASSOCIATED PRESS, (Jan. 15, 2025), <https://apnews.com/article/google-gemini-ai-associated-press-ap-0b57bcf8c80dd406daa9ba916adacafaf>.

existence of the U.K. creative industries but also the public’s ability to enjoy works that would be created by those industries.

The Consultation sets forth several proposed mechanisms to the Proposed TDM Exception in a failed attempt to address rights holders’ concerns—namely conditions of lawful access and express opt-out. While we believe transparency obligations for AI companies are vital, as outlined in our responses to questions 17-22, this measure alone does not alleviate any of our concerns related to the Proposed TDM Exception with lawful access and opt-out measures. We detail below how these so-called “safeguards” fail to address the U.K. Government’s objectives and rights holders’ concerns.

Lawful Access

The Consultation proposes a requirement that those seeking to use the Proposed TDM Exception must have “lawful access” to the copyrighted works they use for AI training. This “safeguard” falls well short of the Government’s objectives as the measure introduces great uncertainty as to its legal interpretation and practicality.

Under copyright law, rights holders may control the circumstances under which their works are made available to others.⁴ Rights holders provide legal access to creative works through licensing terms that specify allowable uses of the works. Under a license, a licensee may be granted legal access for certain uses but not others. If the licensee makes use of a work that is not within the scope of a license, that is considered to be an unauthorized use. Therefore, permitting the use of copyrighted works for AI training without authorization of the rights holder, as has been proposed in this Consultation, is inherently unlawful access.

Importantly, the mere fact that copyrighted works can be *accessed* from lawful services does not also mean that such works can be *used* for the training of AI models without authorization from

⁴ In fact, as signatories to the World Intellectual Property Organization’s (WIPO) Copyright Treaty and Performances and Phonograms Treaty, the UK and US recognize an authors’ exclusive right to authorize “any communication to the public of their works, by wire or wireless means, including the making available to the public of their works . . .” WIPO Copyright Treaty art. 8, Dec. 20, 1996, 36 I.L.M. 65 (1997); WIPO Performances and Phonograms Treaty arts. 10, 14, Dec. 20, 1996, 36 I.L.M. 76 (1997).

the rights holder. The term “lawful access” is misleading in that it suggests that lawful access confers lawful use, which is incorrect.

By way of example, a variety of digital platforms and services make creative works accessible through personal and enterprise subscriptions, through memberships, and other business models. Under the approach proposed in the Consultation, where a rights holder has not executed an “opt-out,” a user could claim to “lawfully access” creative works at a personal subscription level and then turn around and use those works for TDM purposes for a commercial-scale enterprise. But in most cases under current copyright law, this would violate the terms of use governing the personal subscription, making the access unlawful. This undermines the workability of the lawful access requirement and the traditional flexibility and nuance offered by copyright law to rights holders to create tailored licensing models. It will not only fail to bolster rights holders’ ability to control their works, but it will also further restrict access to copyrighted works for AI companies and to the public by incentivizing copyright owners to sequester creative works behind paywalls.

Even where a rights holder has “opted out” (a framework which is severely deficient for a variety of reasons as described below), a “lawful access” requirement may introduce additional questions and uncertainty. For example, when pirated copies of copyrighted works are uploaded to publicly accessible websites or social media platforms or legal copies are illegally uploaded to publicly accessible websites or social media platforms, those copies can be and are scraped by AI companies for AI training purposes. It is unclear whether the “legal access” requirement as used in this Consultation is intended to exonerate AI companies for such scraping and use of these publicly accessible sites and platforms, when there can be no doubt that they should not be.

Regardless of how or whether the uncertainty surrounding the “legal access” requirement can be resolved, an unintended consequence of the requirement will be that rights holders will be forced to increasingly restrict access by placing copyrighted works behind a paywall. This would considerably reduce the availability of high-quality content to the public, which is not in anyone’s interest.

Opt-Out

We oppose so-called opt-out provisions, because they subvert the fundamental nature of copyright and would be unworkable in practice.⁵

If the U.K. Government wants to fulfill the goal of truly supporting rights holders' control of their creative works and ability to be remunerated for such uses, it need look no further than to current copyright law. As an opt-in regime, copyright law properly allows the rights holder to decide whether they want to license their work to an AI company (or anyone else) and on what terms. In contrast, an opt-out regime inappropriately places the burden on rights holders and interferes with their right and ability to license (or choose not to license) their works for AI training purposes. It provides nothing more than an illusion of control, especially given the inability of rights holders to know who is using or trying to use their works.

Moreover, an opt-out regime is limited by virtue of its binary nature for a narrow use; either the work can be used for TDM or it cannot. At present, under an opt-out regime, there is no opportunity for the parties to negotiate terms of use unless the rights holder and AI company reach some further agreement for additional use of the creative work—something they can already do under current copyright laws. AI companies do not need a copyright law exception to do this.

In addition, for a technology where the development and evolution has proven to be quicker than enacting new law and regulations, a legal exception that is technology specific risks becoming quickly irrelevant. Copyright law, on the other hand, provides flexible control for rights holders to respond to new marketplace and technological developments. It achieves the U.K. Government's dual objectives of control for rights holders and access for AI companies.

⁵ See Ed Newton-Rex, *The Insurmountable Problems With Generative AI Opt-Outs* (Nov. 2024), <https://ed.newtonrex.com/optouts>. See also Comments of the Copyright Alliance, Artificial Intelligence and Copyright at 69-71, <https://copyrightalliance.org/wp-content/uploads/2023/10/Copyright-Alliance-Response-to-USCO-NOI-FINAL.pdf>. As we make clear in submission, we oppose any opt-out approach to generative AI training. Even though we oppose an opt-out approach, that does not mean that technologies to prevent scraping and ingestion of copyrighted works do not have a role to play and should not be developed and used.

Importantly, an opt-out regime also conditions rights holders' copyright on effectuating the opt-out. The rights holder would have to comply with the affirmative act of opting out, which, if not properly executed, would strip them of their ability to enjoy and enforce their rights. This imposes a copyright formality that violates Article 5 of the Berne Convention, which states that copyrights are to be enjoyed and exercised by authors without being subject to any formality.⁶ Rights holders must not be required to take affirmative action to prevent others from using their works or have their rights made contingent on the successful execution of opting out.

Independent creators and artists, in particular, are at a disadvantage when it comes to using technological solutions and monitoring for theft of their works. Most do not have the resources or technical expertise to regularly monitor for theft of their works or to take technical and other steps to prevent piracy. By expecting them to now also monitor for AI ingestion of their works and to use technological solutions to prevent ingestion, independent creators and artists would be more burdened than ever by these additional obstacles to the enjoyment of their copyrights that diverts them away from creating new works for the public to enjoy.

Additionally, the opt-out scheme fails to consider the practical difficulties of implementation and thus also fails to achieve the U.K. Government's objective of increased control for rights holders. For example: (1) many copyrighted works have likely already been copied and used for training prior to the opt-out; and (2) despite opting out, copies of the copyrighted works may still be included in the datasets through other means, such as when copies are scraped from other sources such as a licensee of the copyright owner or from a third-party platform where a copy has been posted. The practical effects of opt-out, particularly with regards to works already used to train AI, are also negligible given that it is challenging to remove entire works at scale from an AI model.

While the Consultation mentions that existing technical solutions may assist with opt-out, these tools typically have significant limitations because they are only effective to the extent opt-out is

⁶ Berne Convention for the Protection of Literary and Artistic Works, art. 5(2), Sept. 9, 1886, as revised at Stockholm July 14, 1967, 828 U.N.T.S. 221, 231-32.

recognized and respected, and because these tools are often not designed to be targeted to address scraping for GAI ingestion.⁷ Copyrighted works also often exist as multiple copies on the internet that make it near impossible for a rights holder to enforce the opt-out indicator for every copy of that work. For example, a single song can be streamed on a digital streaming platform, played as the background music of a user uploaded video on a social media platform or in advertisements, or displayed as notes or lyrics on a website. It is impossible for the rights holder to successfully opt out in a way where every single downstream use would be properly tagged with the proper opt-out signal to prevent AI scraping and use. Moreover, copies of works that are available on pirate sites are even further removed from the copyright owner's control, and it is well-known that AI companies have used pirated copies of creative works to train their AI models and have even proliferated pirated copies themselves during the AI development process.⁸ An opt-out regime fails to address or ameliorate any of these problems and certainly does not afford the rights holder any semblance of control. For these same reasons, there is currently a high level of uncertainty in the European Union over what constitutes an effective opt-out,⁹ and as time passes this uncertainty is being exploited by AI developers who continue to train on scraped content despite legitimate efforts from copyright owners to opt out. So, in sum, there can be no doubt that opt-out does not and will not work.

Lastly, as U.S. stakeholders, we feel it is vital to provide further clarity regarding U.S. copyright law and the status of TDM in fair use analysis. With the Consultation referring to the U.S. fair use standard in a section concerning the few countries with broad *statutory* TDM exceptions, we

⁷ Robots.txt protocol is one example. While robots.txt does alert scraping tools not to ingest the associated copyrighted work, it has significant limitations because it is only effective to the extent it is recognized and respected, and it was not designed to be targeted to scraping for generative AI ingestion. Robots.txt may also prevent a search engine from indexing the work. A copyright owner may want their work to be scraped for search engine purposes—so they can be found on the internet—but not for AI ingestion. Even if robots.txt is used, it does not attach to the copyrighted work itself but will operate at the URL or website level.

⁸ Kate Knibbs, *Meta Secretly Trained Its AI on a Notorious Piracy Database, Newly Unredacted Court Docs Reveal*, WIRED (Jan. 9, 2025), <https://www.wired.com/story/new-documents-unredacted-meta-copyright-ai-lawsuit/>.

⁹ We can look to the European Union to see that there is confusion over what is considered a proper “machine-readable” format, a question which has been raised by at least one German court. *See* Landgericht Hamburg [Hamburg Regional Court] Sept. 27, 2024, 310 O.22723, *Kneschke v. LAION*, 310 O.22723 (Ger.). *See also* Roy Kaufman, *AI Rights Reservation: Human Readable is Machine Readable — An Interview with Haralambos (“Babis”) Marmanis*, (Feb. 17, 2025), <https://scholarlykitchen.sspnet.org/2025/02/17/ai-rights-reservation-human-readable-is-machine-readable-an-interview-with-haralambos-babis-marmanis/>.

must reemphasize that fair use is a uniquely American concept that is determined by the U.S. *common law* court system. Arguments that unauthorized uses of copyrighted works (including TDM) may be permitted under U.S. copyright law, particularly on fair use grounds, are overgeneralizations of U.S. law, and to date have proven incorrect. In fact, earlier this month in *Thomson Reuters v. Ross*,¹⁰ which is the first U.S. court decision to address the issue of whether ingestion of copyrighted works for AI training purposes qualifies as a fair use, a federal court rejected the defendant's fair use defense because the use was found to be commercial and not transformative (under the first fair use factor) and found to have harmed the potential market for AI training data (under the fourth fair use factor). While this is only the first of what will be many AI cases addressing fair use questions in the United States, we expect that many courts will hold similarly.

Nevertheless, the important takeaway here is that U.S. copyright law requires courts to engage in a *fact-specific* analysis using an established four-factor test. As such, any AI use, must be carefully analyzed in its particular context. Because fair use analysis is *always* a fact-intensive test tailored to a specific situation it is an inappropriate benchmark to justify sweeping new legal changes outside the United States or the creation of new copyright exceptions. Whether ingestion of copyrighted works for AI training purposes qualifies as a fair use is the subject of nearly forty pending lawsuits in the U.S, which will be decided based on over a century's worth of jurisprudence. As such, there can be no broad characterizations of what AI uses of copyrighted works are allowed in the United States because our laws simply do not provide for *any* such broad exceptions.

¹⁰ See [Thomson Reuters Enterprise Centre GMBH v. Ross Intelligence Inc.](#), 2025 WL 458520, at *6-11 (D.Del. Feb. 11, 2025).

Section C.1 Our proposed approach: Exception with rights reservation

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

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

We do not support the Proposed TDM Exception. However, we do support measures that will increase transparency of copyrighted works used to train AI models (as detailed further below in our responses to questions 17-22). Adequate and appropriate transparency and record-keeping requirements benefit both copyright owners and AI developers in resolving questions regarding infringement and compliance with licensing terms. Transparency obligations also have many other benefits unrelated to copyright. Consequently, transparency by businesses that offer generative AI systems to the public is a crucial component of any AI policy.

Additionally, it is vital for AI companies to implement safeguards to prevent AI models from generating output that is substantially similar to ingested copyrighted works. Overfitting¹¹ and allowing prompts that call for copyright protected-material and “in the style of” are more likely to result in AI-generated outputs that infringe one or more copyrighted works. It is essential that AI companies implement effective safeguards to prevent the likelihood of output-related infringements. This is yet another reason why AI companies should voluntarily license ingested works because when they do so, the parties can negotiate these safeguards.

¹¹ In some instances, AI tools exhibit a machine learning flaw known as overfitting where the output of the system closely matches its training data. See *What is Overfitting?*, IBM, <https://www.ibm.com/topics/overfitting> (last visited Feb. 24, 2025). Overfitting can occur for many reasons, including training the AI model for too long on a limited amount of ingested material or when the ingested material contains large amounts of irrelevant information (also known as “noisy data”). See *What Is Overfitting?*, AMAZON WEB SERVS., <https://aws.amazon.com/what-is/overfitting> (last visited Feb. 24, 2025).

Section C.4 Transparency

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

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

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

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

Yes. Setting aside the other proposals in this Consultation, it is crucial for the Government to adopt regulatory transparency obligations so that AI developers can be held accountable for their unauthorized use of copyright-protected works. Developers of AI models that are made available directly or indirectly to the public that ingest copyrighted works owned by third parties without a license should be required to satisfy transparency standards related to the collection, retention, and disclosure of the copyrighted works they use to train AI.

Adequate transparency regarding ingestion of copyrighted works goes a long way in helping to ensure that copyright owners' rights are respected. Best practices from corporations, research institutions, governments, and other organizations that encourage transparency around AI ingestion already exist, and they enable users of AI systems or those affected by its outputs to know the provenance of those outputs.¹² There is no reason these same responsibilities should not also apply to the ingestion of copyrighted works. However, it's also important to note that there is a big difference between voluntary best practices and binding legal requirements, which is why we support the imposition of legal obligations related to transparency and record keeping.

It is vital that AI developers be legally required to maintain adequate records of what copyrighted works were used to train the AI (and how those works are used) and to make those records publicly accessible and searchable as appropriate, subject to two important exceptions. First, this obligation should not apply to any ingested works of which the AI developer is also the

¹² See, e.g., CONTENT AUTHENTICITY INITIATIVE, <https://contentauthenticity.org/> (last visited Feb. 24, 2025).

copyright owner.¹³ And, second, where there is a license between the AI developer and the copyright owner(s) of the works ingested that authorizes such ingestion for AI development purposes, this obligation should be subject to whatever reasonable confidentiality provisions those parties have negotiated in that license.¹⁴

Where copies of the works ingested are retained, records should also indicate how long copies are retained and what security measures are in place to prevent the copies from being leaked through a cyberattack or otherwise or inadvertently disclosed. Caution in the manner of disclosure should be exercised so that these public disclosures do not further propagate the spread or use of unlicensed copyrighted works.

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

AI companies that ingest and train on copyrighted works that they neither own nor license should be required to identify and list the crawlers they are using to ingest works, and their specific purposes (i.e. whether for indexing or training AI). Such information should be made publicly available and accessible for rights holders. This would give rights holders more information on who is visiting their websites and how to manage that traffic. Ongoing efforts to update existing tools to prevent unwanted web-crawling and scraping such as robots.txt and other “do-not-train” protocols are insufficient on their own. Though we hope and anticipate that new tools will be created in the future to more effectively enable rights holders to signal that they do not want their online works to be scraped and used for AI training purposes, these tools will continue to be insufficient as the sole resolution to resolving outstanding copyright infringement issues and rights holders’ concerns.

Moreover, the protection against stripping of metadata or Copyright Management Information (CMI) and the circumvention of technological protection measures (TPMs) that control access to

¹³ Unless contrary to obligations under other laws, contracts, or collective bargaining agreements.

¹⁴ Data laundering is a major issue in the AI context. The practice of data laundering is an attempt to avoid accountability.

copyrighted works remains an important issue for rights holders. The U.K. Government should, to the extent its Copyright Act does not do so already, consider measures or clarify that AI companies cannot strip metadata and CMI from copyrighted works or circumvent TPMs for AI ingestion, training, and development purposes without authorization from the rights holder.

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

Transparency and recordkeeping should not be onerous or expensive. It simply requires keeping track of and disclosing what datasets are used and, if the AI company creates their own datasets, what sources they used. While the volume of material used in training is often large, systems are not trained on unorganized raw files. Because the datasets are organized and cleaned before training, a commercial market already exists to help AI developers keep such records.

Recordkeeping and disclosure costs are simply a cost of doing business that is necessary in order to promote safe, responsible, respectful, ethical, and unbiased AI systems. These costs must be borne by developers of AI models who are neither owners nor licensees of the copyrighted works at issue.

The cost to rights holders of not imposing a disclosure and record-keeping system is enormous.

Section D.7 Other emerging issues

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

GAI technologies continue to rapidly evolve and scale across the globe at a pace that will undoubtedly pose even greater questions ahead. But it is crucial now more than ever for the U.K. and likeminded countries such as the U.S. to commit to the tried-and-true copyright frameworks which has not only historically weathered multiple technological challenges but also formed the basis for the continued development of those technologies. As the global AI race continues, there will continue to be challenges to find shortcuts and efficiencies in the name of progress. If those shortcuts are utilized without penalties, it will cause a global race to the bottom. We have already

seen challenges come up in the context of newer GAI technologies being developed in China where some AI companies are investigating whether their IP has been misappropriated to develop those competing products.¹⁵ The irony is that this is not unlike the current debate that is occurring over whether the appropriation of copyrighted works to train those very AI models should be done for free on the backs of billions of human creators, copyright owners, and rights holders. It is for this very reason why it is crucial for the U.K. and the U.S. to commit to an approach to AI that is principled, responsible, and ethical—one that respects, protects, and promotes copyright laws that have sustained human creativity and innovation.

Conclusion

We would once again like to thank the U.K. Government for giving us the opportunity to submit these comments. We urge the U.K. Government to rescind the Proposed TDM Exception and continue committing to strong copyright laws so that the flourishing of human creativity that has resulted from the U.S. and U.K.'s long, rich, and mutually supportive creative industries can only continue to grow.

Respectfully Submitted,



Keith Kupferschmid
CEO
Copyright Alliance
1331 F Street, NW, Suite 950
Washington, D.C., 20004

¹⁵ See Cade Metz, *OpenAI Says Deepseek May Have Improperly Harvested Its Data*, The New York Times (Jan. 29, 2025), <https://www.nytimes.com/2025/01/29/technology/openai-deepseek-data-harvest.html>; Daniel Tencer, *OpenAI Valued at \$157bn and Facing Multiple Copyright Infringement Lawsuit, Says China's Deepseek May Have Used Its Data to Train Rival AI Model Without Permission*, MUSIC BUSINESS WORLDWIDE (Jan. 29, 2025), <https://www.musicbusinessworldwide.com/openai-valued-at-157bn-and-facing-multiple-copyright-lawsuits-says-chinas-deepseek-may-have-used-its-data-without-permission/>.