BSA's response to the UK IPO's Copyright and AI Consultation

25 February 2025

Dear Sir / Madam,

The Business Software Alliance (BSA)[1] is the leading advocate for the global software industry. Its members are among the world's most innovative companies, creating software solutions that help businesses of all sizes in every part of the economy to modernize and grow. BSA members are on the leading edge of providing AI-enabled products and services. As such, they have unique insights into the technology's potential to spur digital transformation and practices that can best support the responsible development and use of AI.

BSA respectfully submits the following comments and answers to specific questions from the consultation. As a general matter, BSA would like to highlight that proposed option 2 – a broad data mining exception – best advances the objectives of the AI Opportunities Action Plan and can be implemented without prejudicing the value of the expressive content of copyrighted works.

Essential to AI development is access to a sufficiently large set of data so that computational analysis can reliably identify correlations, patterns, and other metadata to develop an AI model that can make predictions while minimizing risk of inaccuracy or bias. Some of that data may be a component of a copyrighted work, but the use of the data for AI training is normally not related to the work's creative expression.

[1] *BSA's members include: Adobe, Akamai, Alteryx, Amadeus, Asana, Atlassian, Autodesk, Bentley Systems, Box, Cisco, Cloudflare, Cohere, Dassault Systemes, Databricks, DocuSign, Dropbox, Elastic, EY, Graphisoft, HubSpot, IBM, Informatica, Kyndryl, MathWorks, Microsoft, Notion, Okta, OpenAI, Oracle, PagerDuty, Palo Alto Networks, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Trend Micro, TriNet, Twilio, Workday, Zendesk, and Zoom Communications, Inc.*

Option 2 would give AI developers confidence to train AI models and systems in the UK. At the same time, BSA encourages more work on consensus-based, machine-readable tools to indicate that a rights-owner does not want a website used for training purposes. In fact, there is already a burgeoning ecosystem for such private collaboration between AI developers and deployers on the one hand, and licensors of creative content on the other.

By contrast, proposed option 3 focuses on measures that benefit licensors of commercial content and would come at the expense of AI-driven science, ingenuity, and creativity in the United Kingdom (UK). Contrary to the objectives of the AI Opportunities Action Plan, this proposal would hinder the development and deployment of AI in the UK and harm the UK's AI-driven growth potential and global competitiveness.

We urge the UK IPO not to adopt proposed option 3. It would stifle not only the AI ecosystem, but also creative and innovative endeavor, scientific research, and the generation of new IP in the UK more broadly. Rather than adopting a limiting framework that gives large copyright licensors the ability to impede and stifle progress in AI training, we urge the UK IPO to foster an approach that is better suited to realizing AI's significant social, economic, and scientific potential.

In our responses to specific questions below, we focus on the importance of text and data mining (TDM) technologies for the development of AI in the UK, on the state of the law as well as on text and data mining rules applicable outside of the UK.

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

BSA would like to stress that Option 3: "A data mining exception which allows right holders to reserve their rights, underpinned by supporting measures on transparency", could only meet the government's objective if it is scoped in a manner that enables compliance with the opt out, and provided that it is clarified that **the reservation of rights does not function as a right to prevent text and data mining**.

In any case, the reservation of rights under Option 3 should meet standards of machine-readability and interoperability. We urge the government to link any compliance with the reservation of rights to a standard that is representative of industry conversations, which continue to evolve with new use cases and technological developments. This allows for global operability, feasibility and legal certainty for the benefit of all stakeholders.

Furthermore, while we support the ability for rightsholders to express their preferences as to whether their works are used in AI training, as provided by Article 4 of the EU

Copyright Directive, we do not agree that an opt out should have the legal effect to prevent all text and data mining as suggested in paragraphs 77-80 of the UK IPO's consultation. This would go beyond what is implemented in the EU, where a rightsholder is able to reserve their rights to prevent the TDM exception from applying, but allows for other exceptions and defences to apply where relevant, hereby allowing activities such as inferencing and other forms of data analysis.

By contrast, the proposed opt out in option 3 risks creating a right to prevent data mining, while overlooking other exceptions that may apply. It also swings the pendulum in favour of rightsholders versus balancing the interests that would be beneficial to greater society.

Requiring licences to analyse data within copyright works would provide rightsholders with unprecedented and limitless control over downstream innovation, blocking open source AI development, severely stifling innovation in the UK and distorting competition.

Therefore, we would recommend the UK government to introduce an opt out that is more in line with the EU by enabling rightsholders to indicate their preferences regarding AI training using standardised technical means, without creating a right to prevent training.

Question 2. Which option do you prefer and why?

We believe that option 2 presents the best environment for AI development and supports AI advancement for greater society. Expanding the TDM exception to commercial uses would enable businesses to mine copyrighted materials without needing permission, provided they have lawful access. This could significantly boost AI development, innovation, and global competitiveness by reducing licensing barriers and attracting tech investment. SMEs would particularly benefit from lower costs and streamlined data access, while industries like finance, healthcare, and journalism could leverage large scale analysis for better products, services, and decision-making.

The government's preferred option 3 would not meet the government's objective of enabling access to data, if it is implemented with an opt out that has the legal effect of creating a new right to prevent data mining as currently proposed. It is not in line with the EU TDM exception. This type of opt out would impact the ability to analyse information that is contained within publicly available works, since data mining is fundamental to data analysis.

We understand that rightsholders want to manage how their online content is used. We are supportive of controls exercised via machine-readable opt outs through harmonized industry standards such as the C2PA. Machine-readable reservations expressed via

signals that can be consumed by crawler bots at the point of access or collection not only take into account how data ingestion on the web works but are optimal. For rightsholders, the machine-readable signal provides a mechanism to effectively express a reservation; for AI developers, it provides certainty that if they have lawful access to the content, they will not face legal claims after the fact. There are ongoing conversations within and among industries to continuously improve these controls.

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

We agree that the text and data mining exception should apply broadly to both commercial and non-commercial uses. However, text and data mining activities conducted for research purposes should not be subject to the opt out, otherwise this would erode productive and existing research that relies on the current TDM framework. To support and develop AI innovation - we need to ensure developers and researchers are able to use publicly-available data. This will yield huge advances in fields from physics to medicine to education, continuing economic growth, and tens or hundreds of millions of new jobs over the next several years.

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

As mentioned above, the opt out should not have the legal effect of preventing data mining. Therefore, we disagree with paragraphs 77-80. Instead of requiring the opt out to have legal effect, the opt out can be implemented in standards that indicate that the work should not be used to train AI models. This approach is flexible and can accommodate new technologies and uses.

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

Copyright holders always retain their right to seek legal recourse if their IP rights are infringed. However, in the specific context of reservations from AI training, it is not appropriate to establish special legal consequences relating to opt-outs.

Parallels can be drawn with the longstanding norms surrounding the operation of web crawlers. These norms have effectively governed how web crawlers interact with web pages, without the need for specific legal frameworks to enforce compliance. This model could serve as a precedent for handling opt outs for TDM. Companies will adhere to established standards because doing so fosters trust, ensures good relationships with content creators and avoids reputational harm.

Whether a copyright holder chooses to initiate a legal claim is a separate question, governed by carefully defined legal standards. The determinative standards implicate questions of ownership, validity, scope, and duration of relevant IP rights under applicable statutory and common law standards. As such, any final judicial determination of copyright infringement possesses a clear and objective legitimacy within the legal system.

By contrast, an expressed preference by any given copyright holder to "opt out" of AI training does not possess the same objective legitimacy with regard to the copyright holder's assertions regarding the ownership, validity, scope, or duration of any claimed IP rights. Instead, such an "opt out" preference reflects a subjective and unilateral preference of one person – a preference that has not been tested or evaluated through the rigors of judicial process, let alone endorsed in a judicial judgment.

Conferring legal weight to such a subjective and unilateral preference risks undermining substantive legal standards, and distorting predictable judicial process, in a way that is not consistent with established copyright law or procedure.

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.

We agree that a reservation of rights should be machine readable. This enables AI developers to identify opted out works at scale, allowing for the preferences of rightsholders to be respected efficiently and effectively. However, we do not agree that the effects of the rights reservation should apply more broadly than the individual copy to which it applies. The technical challenges to identifying other copies of a work that do not have a machine-readable rights reservation would be overwhelming, particularly since AI training does not typically involve reviewing content for its expressive meaning. Industry is working on improved technical solutions that may enable an opt out signal to be interpreted more broadly, and the government should allow such technologies to develop.

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

The government should not play a role in commercial agreements between parties for licensing. Furthermore, we do not agree with the suggestion that collective management organisations (CMOs) should have the unilateral right to prevent non-consumptive AI training activities by levelling licensing demands against any entity

engaged in AI training in the context of AI development or deployment. Conferring such a unilateral right on CMOs would increase conflicts with the limited copyright exceptions necessary to foster AI training in the UK, thus greatly weakening any benefits from a TDM exception in the first place.

First, CMOs may adopt overly broad opt outs that cover entire repertoires of works. This would prevent data mining on works even when individual authors or creators have no objection. This would also create a lack of clarity around what has been opted out, since CMOs do not necessarily have clarity over what exactly their repertoire contains.

Second, turning first to the ability for a CMO to opt out - an opt out should be a decision made by the rightsholder. In some cases the rightsholder may want to use an opt out to prevent access to their work, rather than use the opt out as a basis for licensing. Alternatively, a rightsholder may not want to opt out and see value in allowing their works to be analysed in data mining.

Third, even if a rightsholder has reserved their rights, the use of CMOs to provide access to works for AI training creates jurisdictional challenges. This is complicated further given the fact that CMOs typically operate in relation to specific rights in one country, copyright laws vary across jurisdictions and models are trained and used in multiple jurisdictions.

Fourth, it is also the case that many rightsholders will want to negotiate directly with AI developers. There is evidence of many partnerships developing between rightsholders and generative AI companies to enable creators to make money and expand access to their materials. These include partnerships to build and improve AI models, using content that would otherwise be inaccessible, as well as partnerships that would create new information experiences and revenue models, using content in ways that would otherwise not be permitted by copyright. Enabling CMOs to opt out and license works will undermine such private agreements.

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

AI developers disclose information that can be publicly available about their training material through model cards. Outside of that, we support transparency requirements that are thoughtful, measured, but would not require AI developers to give up IP-related and confidential information. In addition, AI developers negotiate terms relating to training material disclosure in their commercial agreements with their downstream customers and this information is not intended to be made public for a reason or protecting business interests on both sides.

Transparency requirements should:

- Avoid excessive disclosure requirements (such as an exhaustive inventory of training data), which could lead to high compliance costs, place undue burden on small and start-up developers and jeopardize trade secrets as well as violate confidential arrangements.
- Remain adaptable, allowing requirements to evolve alongside industry standards. To this end, AI developers should be encouraged to continue to engage in transparency measures in the form of model cards and transparency notes, as well as publishing research. Allowing such transparency mechanisms to evolve organically via voluntary collaboration between AI developers/deployers and licensors of commercial content is preferable to a top-down government mandate that does not account for industry standards and best practices.
- Not serve as a pretext to gather up just enough information to launch legally dubious litigation claims. This is not an appropriate aim for such transparency mechanisms. To design transparency mechanisms in the service of fabricating costly copyright litigation is to impose a tax on innovation for UK companies across numerous sectors, while also interfering with the UK government's objective of encouraging access to data for socially and economically beneficial purposes.

We recognize the desire of copyright holders to understand if model developers are respecting their choice, specifically any reservation of rights. We believe that pairing (i) a general description of the training data source, in a narrative form, with specific references to large collections, along with a description of the purpose of the model; and (ii) information about the crawlers or bots used to collect information, will provide copyright holders with sufficient information to investigate if their choice is being respected.

Question 23. What are your views on the EU's approach to transparency?

We believe that the current draft requirements for technical documentation and information sharing to downstream providers contained in the EU draft Code of Practice are too prescriptive, requiring disclosures of very detailed information, including granular disclosures on training data, and goes beyond the Act's provisions and Annexes XI and XII, implicating concerns about disclosures of trade secrets and proprietary information. Moreover, we note that Annex XII of the AI Act already has detailed requirements and recommend avoiding establishing disproportionate requirements about the sharing of information as they may become outdated quickly and are unlikely to fit different GPAI models equally.

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?

Creating a level playing field for AI development in the UK is best achieved by creating a favourable environment in the UK for AI development. If the UK creates a prohibitive and deterrent licensing framework for AI training that must be complied with, irrespective of where the models are trained, this is likely to impact the models that are available in the UK. BSA urges the government not to introduce any measures that would limit the ability to use models trained outside of the UK in the UK. The government should instead expressly allow for models trained outside of the UK to be deployed and used in the UK.

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

It is important that any reservation of rights associated with a TDM framework does not interfere with existing exceptions, such as the temporary copies exception, which plays a foundational role in enabling innovation across various sectors – from agriculture, automotive, and aerospace to finance and pharmaceuticals. This exception allows innovators to develop and deploy AI tools that draw on a corpus of sufficient size and depth to function effectively.

Question 30. 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.

We do not believe that the computer-generated works provision is needed to protect works generated using AI, since in many cases there will be a human author. AI is a tool like any other that is being used by humans to create. Often, the use of AI to generate a work requires substantial creativity and judgment to produce results from the AI tool. AI tools allow humans to make creative decisions at various points throughout the creative process. Many tools allow users to provide detailed prompts and iteratively generate outputs, including selecting between options and further refining the output that can be fed back into the system. This iterative process demonstrates significant judgement in making creative choices, which exceeds the minimal degree of creativity required to give rise to copyright protection and enable an author to be identified.

If the computer-generated works provision is deleted from the CDPA, it would be useful for guidance to be provided that explains that works created using AI are still subject to copyright protection, provided there is a human author. This clarification will encourage continued use and investment in AI for the benefit of the creative sector.

_____

For more information, please contact Thomas Boué, Director General, EMEA at BSA, via email at thomasb@bsa.org.