Q1. Do you agree that option 3 is most likely to meet the objective set out above?

B. Copyright and Artificial Intelligence

Q1. Do you agree that option 3 is most likely to meet the objectives set out above?

No. There are multiple reasons for rejecting this proposal. They are as follows.

1. High costs and burdensome licensing associated with introducing option 3.

(a) Transactional costs relating to multiple opt-outs: it is possible that a rightsholder may have different opt-out statements and if so, how can different prices for different types of uses be set? It is not clear how this can be achieved or enforced under option 3. On the other contrary, if different statements are not permissible, then how does a rightsholder enforce their rights?[i] The Consultation does not clarify these issues and in the light of the proposed system, option 3 will lead to additional transactional costs relating to multiple outputs.

(b) Option 3 will also lead to **transactional costs in relicensing content that has already been licensed** to organisations with legal access to copyright works. Such a mechanism will lead to burdensome transaction costs, which could be borne by those with organisations with large funds, but not those with recourse to limited funding. Such a system will make the UK uncompetitive in comparison to other countries who do not have such legal obligations

(c) Practical and technical issues. One of the key issues is when the mining should be carried out. Should it be *ex-ante* or *ex-poste*? Both are riddled with their own problems. The former – *ex-ante* – goes against copyright's fundamental principle of opt-in which permits the use of a copyright work, if permitted to do so. The latter – *ex-poste* – is technically not feasible as it would require the removal of mined text or data from the dataset of GenAI applications.[ii]

(d) Issue of "opt-out shopping". Where the same work carries different opt-out statements on different websites, whose responsibility will it be to ensure an accurate implementation of the opt-out and avoid "opt-out shopping"? [iii] The consultation is silent on this matter and does not address how it can be resolved.

(e) Lack of evidence at present that training GenAI algorithms leads to copyright infringement. The current consensus as reflected in the literature is that training GenAI algorithms will not lead to permanent reproduction of the training data. Case law will resolve this matter however, UK should not wait for those outcomes to act in the AI field and proposing option 3 when there are more questions than answers, seems short-sighted.

Q2. Which option do you prefer and why?

We prefer option 2 – a broad data mining exception. The reasons are as follows.

(a) A broad data mining exception has been adopted in countries such as Singapore and Japan with much success. A research report published in 2024 pointed to the fact that countries with broad and flexible exceptions tend to see better progress in the creative, educational and research sectors whilst benefitting from developments in technology in a timely manner.[iv] This type of evidence strengthens the argument for option 2; therefore, there is no reason why a broad data mining exception cannot be introduced in the UK.

(b) Fair use can address data mining, as seen in USA. The Consultation does not back a fair use type exception, but at the same time, fails to say why the fair use approach will be incompatible with a broad data mining exception.

(c) A broad data mining exception will mean that **UK will have access to more data leading to greater predictive capabilities**. A reservation right restricts the amount of data that can be used, stunting the advancement of AI predictive capabilities[v], a broad data mining exception will lead to results which have better accuracy and veracity.

C. Our proposed approach: Exceptions with right reservation

Q3. Do you support the introduction of an exception along the lines outlined above?

No, we do not. As outlined with Q1, we reject the proposal of an exception subject to a reservation right.

Q4. If so, what aspects do you consider to be the most important? If not, what other approach do you propose and how would that achieve the intended balance of objectives?

We recommend a broad data mining exception, complemented by an AI levy as compensation for the creators of literary, dramatic, musical and artistic works.

(a) Broad data mining exception complemented by the payment of a single equitable remuneration. Drawing on the experience of countries such as Germany which impose a levy on high-tech equipment, this proposition recommends a levy to be imposed on users of AI systems that generate creative literary, dramatic, musical and artistic (LDMA) content that compete with the original.[vi] Such a system will serve the purpose of creating a new revenue stream to support creators and authors of creative works. The monies collected can be distributed through Collective Management Organisations (CMOs) and the existing licensing scheme without the need to 'reinvent the wheel'.

(b) This would ensure that **AI system developers will not be disadvantaged** in advancing new AI models whilst **at the same time, human authors of creative LDMA works will be compensated**. This system further ensures alignment with Article 5(1) of the InfoSoc Directive – temporary acts of reproduction – which applies if it does not lead to commercial exploitation. If it does, the exception will be defeated, and remuneration will be applicable. In this manner, an AI system levy will compensate creators when their outputs are commercially exploited.

(c) Introduce a flexible open norm by reforming UK copyright law. As already outlined in Q2, reforming UK copyright law to introduce a flexible copyright exception will lead to better progress in the creative, educational and research sectors whilst benefitting from developments in technology in a timely manner.[vii] There is therefore, no reason why a flexible open norm cannot be introduced in the UK.

Q5. 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.

It will greatly limit high quality research. The accuracy and veracity of AI models will depend on high quality research emanating from academic institutions, amongst others. In view of a reservation right, these scholarly outputs which have already been licensed will need to relicensed making the publishers gatekeepers in the process. This is a negative influence of the reservation right and will prove to be a major drawback for academic institutions.

Q6. What action should a developer take when a reservation has been applied to a copy of a work? Q7. What should be the legal consequences if a reservation is ignored? Q8. Do you agree that rights should be reserved in machine-readable formats? Where possible, please indicated what you anticipate the cost of introducing and / or complying with a rights reservation in machine readable format would be.

We have addressed Q6-Q8 together as we reject the reservation right and therefore these questions are not relevant to our present submission. A reservation right is also not applicable when dealing with 'technical copying' (temporary reproduction) when analysing trends, patterns, facts and other non-protected elements of a work. As such, the reservation right becomes redundant at this stage of the process.

The cost of introducing a reservation right has been addressed under Q1. High transactive costs, burdensome relicensing and practical/technical issues are some of the associated costs.

C.2 Technical Standards

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

Yes, there is indeed a need for greater standardisation of rights reservation protocols. While existing mechanisms such as robots.txt and metadata-based approaches allow rights holders to signal their preferences, these systems lack uniformity and broad adoption by the industry. Many right holders in particular individual creators (such as artists) face difficulties in implementing these measures effectively, as they lack technical expertise and/or resources.

Standardisation would ensure that all AI developers adhere to a consistent, and legally recognised framework for rights reservation. It would also provide AI developers with legal certainty, reducing ambiguity over whether they can use specific content for training AI models.

Question 10: How can compliance with standards be encouraged?

Compliance with rights reservation standards can be encouraged by a combination of the below:

Regulatory incentives and obligations: The government could introduce legal requirements that make compliance with standardised rights reservation protocols mandatory for AI developers. This could include obligations to respect standardised metadata signals and a requirement for AI developers to provide transparent records of how they acquire training data.

Industry self-regulation: Collaboration between AI developers, creative industries, and standards bodies (such as W3C and the Internet Engineering Task Force) could help establish best practices for AI training compliance. AI firms could be incentivised to adopt these standards through certification schemes or regulatory relief. However, it must be noted that industry self-regulation may not be sufficient on its own and must be backed up by government incentives and other approaches listed below.

Technological approaches: AI companies could be required to integrate automated tools that detect and respect machine-readable rights reservations (e.g., through APIs or blockchain-based verification systems).

Educating right holders. To ensure compliance, rights holders, particularly individual creators need to understand how to implement and enforce rights reservations effectively through standardised tools.

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

Yes, the government should play a role in ensuring the effective standardisation and adoption of rights reservation protocols. This role should include:

Legislative and Regulatory Supervision

The government should provide a clear legal framework mandating that AI developers respect rights and reservations and incorporate compliance mechanisms into their data collection practices. Furthermore, to ensure compliance, the Government should introduce penalties for AI companies that ignore standardised rights reservations.

Facilitation of Industry Standards

The government can act as a convening force, bringing together AI developers, content creators, technical bodies and other relevant stakeholders (i.e the users of generative AI services) to agree on universal standards for rights reservation protocols.

International Coordination

International coordination and engagement with global AI governance frameworks is crucial for the UK (e.g., EU's AI Act and the European Copyright Directive) to ensure that UK-based rights reservation standards align with international best practices, preventing forum shopping by AI firms.

The EU AI Act imposes specific obligations on AI system providers, especially those offering general-purpose AI models. Providers are required to establish policies that respect EU copyright laws, including mechanisms to identify and comply with rights reservations. This ensures that AI systems do not infringe upon protected content during their training processes.

The EU's Copyright Directive (Directive 2019/790/EU) introduces exceptions for text and data mining, allowing certain uses of copyrighted materials without explicit permission. However, rights holders can opt out by clearly expressing their reservations. Arguably, this opt-out mechanism allows content creators to limit and control the use of their works in AI training datasets. It must be noted that the EU's

approach is subject to various criticism[1]. Therefore, the UK should assess the applicability of EU copyright rules before establishing its own standards for AI dataset training.

C. 3 Contracts and Licensing

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 licensing practices for AI training do not fully meet the needs of creators and performers. The reasons for this are set out below.

Lack of Transparency

Many creators do not know whether their works have been used for AI training, as AI developers rarely disclose their data sources. This limits creators' ability to negotiate fair compensation.

Limited Control Over Licensing

While licensing is often done via collective management organisations (CMOs), many creators feel that licensing terms are dictated by industry norms rather than their preferences.

Imbalance in Negotiation Power

Many AI training datasets rely on broad scraping of internet content, bypassing traditional licensing structures. Creators often lack the bargaining power to negotiate fair deals.

Industry Expectations to License Works for AI Training

[1] For instance see Kaigeng Li, Hong Wu, Yupeng Dong,'Copyright protection during the training stage of generative AI: Industry-oriented U.S. law, rights-oriented EU law, and fair remuneration rights for generative AI training under the UN's international governance regime for AI' Computer Law & Security Review, Volume 55, 2024,

Some performers and creators feel pressured to accept AI training clauses in their contracts, even if they would prefer to opt out. Academic publishing is a good case in point. Many publishing contracts include a very broad grant of rights–to the publisher for them to exploit those rights and to license the rights to third parties. Authors have very limited power to amend/alter these terms. In such cases, when the publisher becomes the de facto copyright holder of a work, it is very difficult for authors to stake a copyright claim when their works are being sold to train AI.

Question 13: Where possible, please indicate the revenue/cost that you or your organisation receives/pays per year for this licensing under current practice.

Not applicable to our organisation.

Question 14: Should measures be introduced to support good licensing practices?

Yes, stronger measures should be introduced to ensure licensing for AI training is fair, transparent, and enforceable. Possible measures include:

Enhanced Transparency Requirements

AI developers should be required to disclose their training datasets so that right holders know when their works are used. In this respect, the UK could potentially follow the EU AI Act's transparency rules, which mandate public summaries of training data.

Standardised Licensing Frameworks

The government should develop a clear licensing framework for AI training, ensuring that licensing terms are fair and enforceable. This could include a model contract template for licensing AI training data, as well as recommended pricing structures.

Stronger Collective Licensing Mechanisms

CMOs should be empowered to negotiate AI training licences on behalf of creators, ensuring fair remuneration.

Government Oversight and Dispute Resolution

A regulatory body such as the Intellectual Property Office can oversee AI licensing, ensuring compliance. It can also address disputes between AI firms and rights holders. Furthermore, standard licensing audits should be conducted to track whether AI firms comply with licensing agreements.

Exploring Rights Reservation Mechanisms

Arguably, to ensure that the rights of copyright holders are fully protected, AI training should be subject to opt-in rather than opt-out licensing, ensuring that content is only used with explicit permission of the right holders.

Q15) Should the government have a role in encouraging collective licensing and/or data aggregation services? If so, what role should it play?

Yes, the government should play an active role in encouraging collective licensing and data aggregation services. To facilitate this, the Government should

- Ensure that CMOs are empowered and have the necessary skill set to negotiate fair licensing terms for AI training.
- Provide training to rights holders so that they are aware of their rights.
- Provide regulatory guidance on standardised AI licensing frameworks, helping AI developers and CMOs establish fair agreements.
- Fund the development of digital licensing platforms, like initiatives in the EU AI Act, to automate and simplify licensing processes.
- Encourage interoperability between UK and EU licensing frameworks, ensuring that UK creators can benefit from international licensing opportunities.

Q16) Are you aware of any individuals or bodies with specific licensing needs that should be taken into account?

Apart from the creative industries, organisations in specific sectors may have unique licensing requirements.

The Education Sector and Relevant Bodies

AI-generated educational tools require access to vast amounts of licensed content, raising concerns about fair compensation for textbook authors and academic publishers. As a result, certain educational bodies may have specific licensing needs. For instance, Universities UK (UUK) represents higher education institutions and plays a role in shaping intellectual property policies related to AI use in academic research. Additionally, Jisc provides digital solutions for UK education and research, including licensing models for AI-driven academic databases.

News and Media Organizations and Relevant Bodies

AI models frequently train on news articles and journalistic content, yet publishers often receive little to no compensation. Therefore, some organisations in this sector may have particular licensing needs. For example, the News Media Association (NMA), which represents national and regional news publishers, advocates for AI licensing agreements to prevent unauthorized text and data mining. Similarly**,** Reuters and the BBC are actively involved in developing licensing models for AI use of journalistic content.

AI Developers, Startups, and Relevant Bodies

AI firms, particularly startups and SMEs, often struggle to navigate the fragmented licensing landscape. For instance, TechUK, which represents the UK technology industry, advocates for clear AI licensing policies to ensure the legal and ethical use of training data.

Other Relevant Bodies

The British Library and National Archives manage large public datasets and play a key role in data licensing frameworks for AI use. As a result, they may also have specific licensing needs.

Transparency

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

Yes, BILETA agrees that AI developers should disclose the sources of their training material, particularly when it comes to copyrighted works. This transparency is crucial for several reasons:

- **Accountability and Trust:** Disclosing the sources of training data allows creators, users, and the public to understand how AI models are developed and to assess the potential impact on copyright-protected works. This fosters trust and accountability in the AI development process.
- **Copyright Compliance:** Transparency enables the verification of whether copyrighted works have been used with proper authorization and licensing, ensuring compliance with copyright laws and protecting creators' rights.
- **Fair Use/Fair Dealing Analysis:** Knowing the sources of training data is essential for determining whether the use of copyrighted material falls under fair use/fair dealing exceptions. This analysis often hinges on factors like the purpose and character of the use, the nature of the copyrighted work, and the impact on the potential market for the original work.
- **GDPR Principles:** Transparency in AI training aligns with the principles of the General Data Protection Regulation (GDPR), particularly Articles 5 and 12, which emphasize the importance of lawful, fair, and transparent data processing. Disclosing training data sources helps ensure that AI development respects data protection rights.

While protecting legitimate trade secrets is important (Recital 63 GDPR), disclosing the specific sources of copyrighted training data is essential for transparency and accountability. This approach is supported by the Court of Justice of the European Union (CJEU) in the *C-634/21 SCHUFA Holding and Others (Scoring)* case, where the Advocate General emphasized the importance of transparency and explainability in automated decision-making systems, including those using AI (see [AG 54, 57, 58]).

The Role of the EDPS and Explainable AI (XAI)

The European Data Protection Supervisor (EDPS) echoes the need for transparency, stating that AI tools should go beyond simple classifications to provide nuanced explanations of the underlying logic (see the EDPS's report on "Deepfake

Detection" - https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/deepfake-detection_en). For copyright analysis, this includes details about the copyrighted sources used in training, the analysis methods employed, and the key indicators that identify potential copyright infringement. This transparency fosters accountability, trust, and ethical use, ensuring compliance with data minimization (Article 5(1)(c) GDPR), purpose limitation (Article 5(1)(b) GDPR), and storage limitation (Article 5(1)(e) GDPR) principles.

Explainable AI (XAI) plays a critical role here (see the EDPS's "TechDispatch" on "Explainable Artificial Intelligence" - https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2023-11-16-techdispatch-22023-explainable-artificial-intelligence_en). By revealing how AI systems arrive at their conclusions, XAI can help determine if a work incorporates substantial elements of copyrighted material or if it constitutes a transformative use. It can pinpoint specific features and provide human-readable explanations of the AI's reasoning, aiding in copyright analysis. XAI techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) can identify the most influential features in the AI's output, further assisting in understanding the impact of copyrighted training data on the final creation (see https://arxiv.org/abs/2305.02012).

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

AI firms should disclose training data sources with a level of granularity that balances protecting creators' rights with fostering innovation. This means clearly identifying the copyrighted works used in training, including titles, authors, and any relevant licensing information. This approach aligns with the UN's Resolution A/78/L.49 on Safe, Secure, and Trustworthy AI, which underscores the need to protect intellectual property rights, particularly copyright, in the context of AI development (see https://docs.un.org/en/A/78/L.49). For example, Adobe's Content Credentials (CR) allows for proper attribution, assessment of potential infringement, and compliance with licensing agreements (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).

Highlighting Significant Contributions and Data Usage

If specific copyrighted works play a significant role in shaping the AI's output, this should be highlighted, potentially by indicating their weight or influence in the training process, especially if they contribute to distinctive features in the generated content (see AG opinion in *C-634/21 SCHUFA Holding and Others (Scoring)* [AG 58]). AI firms should also explain how the copyrighted data was used in training, including any modifications or transformations applied, aligning with the EDPS's emphasis on nuanced explanations of the underlying logic of AI systems (see https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/deepfake-detection_en). This helps assess the extent to which the training data influenced the final output and whether it constitutes fair use/fair dealing or transformative use.

Protecting Trade Secrets and Ensuring Accessibility

However, this level of granularity should not compromise legitimate trade secrets or confidential information related to the AI model's architecture or training process, echoing the GDPR's acknowledgment of the need to protect trade secrets (Recital 63 GDPR) – see [AG 54] in *C-634/21 SCHUFA*. The disclosure should be clear and concise, avoiding unnecessary complexity or technical jargon that might hinder understanding, ensuring that the information is accessible to creators, users, and other stakeholders in line with the principle of accessibility of the European Convention on Human Rights (Article 8(2) and Article 10(2)) and European Court of Human Rights case law (see, e.g., *Kennedy v the United Kingdom* App no 26839/05 (2010) 52 EHRR 247).

24. What transparency should be required in relation to web crawlers? Please give us your views.

Web crawlers play a vital role in the digital ecosystem, but their operation must be transparent and accountable to ensure responsible data collection practices. This response outlines key principles for achieving transparency in web crawling, drawing on relevant legal frameworks and industry best practices.

Identification and Purpose

Web crawlers should be clearly identifiable, with readily accessible contact information for the operator. This allows website owners and users to understand who is collecting data and how to contact them with any questions or concerns. The purpose of the web crawling activity should be clearly stated, including what data is being collected and how it will be used. This enables website owners and users to assess the legitimacy and potential impact of the data collection.

Respect for Website Policies and User Privacy

Web crawlers must respect website owners' preferences and user privacy. This includes strict adherence to the robots.txt file on websites, which specifies which parts of the site should not be crawled. Crawlers should be designed with privacy in mind, minimizing the collection of personal data and ensuring compliance with GDPR principles like data minimization (Article 5(1)(c)), purpose limitation (Article 5(1)(b)), and storage limitation (Article 5(1)(e)). This also entails providing clear privacy policies and obtaining informed, freely given, specific, and unambiguous consent when necessary, as stipulated by Articles 13 and 14 of the GDPR and relevant CJEU case law, such as *C-673/17 Planet49* and *C-61/19 Orange Romania*.

Data Handling and Security

Robust security measures are essential to protect the collected data from unauthorized access, use, or disclosure. Clear data retention policies should specify how long the data will be stored and how it will be disposed of securely.

Transparency in AI Training

If the crawled data is used for training AI models, this should be disclosed, along with information about how the data is used and any potential impact on the generated outputs. This aligns with the principles of explainability and accountability in AI, as highlighted in the CJEU's *C-634/21 SCHUFA* ruling.

Copyright Compliance

If the crawled data includes copyrighted material, proper attribution and licensing should be ensured, respecting creators' rights and complying with copyright laws. This echoes the principles promoted by the Coalition for Content Provenance and Authenticity (C2PA) and Adobe's Content Credentials (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).

Accountability and Redress

Web crawler operators should provide clear mechanisms for website owners and users to raise concerns or complaints about data collection practices. This could include contact information, feedback forms, or dispute resolution processes. Establishing accountability frameworks, potentially through industry standards or regulatory bodies, can help ensure that web crawlers operate responsibly and transparently.

Addressing the Challenges of Data Scraping

The OECD highlights the diverse nature of data scraping and the associated legal challenges, emphasizing the need for clear policies and standardized tools to protect IP rights (see https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf). A collaborative, international effort is needed to develop a code of conduct, standardized contracts and tools, and initiatives to educate stakeholders on ethical data usage (see https://c2pa.org/principles/).

How much can a web crawler do under current law on a website?

In the US web crawlers, using principles from UK law, could be considered a trespass to chattels in certain circumstances -see Ticketmaster v Tickets.com case No. 99-CV-07654 U.S. District Court, Central District of California March 27, 2000. However, this needs to be compared to eBay v Bidders Edge 100 F. Supp. 2d 1058 (N.D. Cal. 2000), Web crawlers, per these cases, could slow down websites and redirect traffic, causing monetary loss. If that is sufficient, there might be an action for trespass to chattels possible.

25. What is a proportionate approach to ensuring appropriate transparency? Please give us your views.

(Please refer to question 23 above)

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

Estimating the precise costs of introducing transparency measures for AI developers is challenging, as it depends on various factors such as the specific measures implemented, the complexity of the AI models, and the existing infrastructure of the developers. However, we can analyse potential cost categories and consider insights from relevant sources.

Potential Cost Categories

Potential cost categories include:

- **Data documentation and tracking:** Meticulous documentation and tracking of training data, including provenance, licensing, and usage rights, could involve significant initial investment and ongoing maintenance costs.
- **Development of Explainable AI (XAI) tools:** Research and development costs might be incurred, along with potential costs for training staff on using and interpreting XAI outputs.
- **Adoption of transparency tools:** Integrating tools like Content Credentials or blockchain-based provenance tracking could involve costs for software licenses, implementation, and ongoing maintenance (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).
- **Compliance with legal frameworks:** Ensuring compliance with evolving legal frameworks and regulations related to AI transparency, such as the EU AI Act, might require legal consultations, audits, and adjustments to AI development processes.

- **User interfaces and communication:** Developing user-friendly interfaces and communication materials to explain AI transparency measures to users could involve design and development costs.

Insights from Regulators and Organizations

The CNIL (French data protection authority) recommends that AI developers adapt their approach to fulfilling data subject rights requests according to costs and practical constraints, acknowledging that cost considerations are relevant in implementing transparency measures (see https://www.cnil.fr/en/ai-and-gdpr-cnil-publishes-new-recommendations-support-responsible-innovation#:~:text=In%20some%20cases%2C%20the%20cost,possible%20protection%20of%20individuals'%20rights). The EDPS (European Data Protection Supervisor) emphasizes the importance of transparency and explainability in AI (see https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2023-11-16-techdispatch-22023-explainable-artificial-intelligence_en), but also acknowledges the need to balance this with the protection of trade secrets and intellectual property rights, suggesting that cost-effective solutions are needed.

Collaboration and Open-Source Solutions

Collaboration between AI developers, industry consortia, and governments could help share the costs of developing and implementing transparency measures. Utilizing open-source tools and standards for transparency could help reduce costs for individual AI developers (see, e.g., https://c2pa.org/). While transparency measures may involve upfront costs, they can lead to long-term benefits, such as increased trust, reduced legal risks, and improved accountability, potentially outweighing the initial investment.

Introducing transparency measures on AI developers will likely involve costs across various categories, but the specific amounts will vary depending on individual circumstances. Governments and industry stakeholders should consider cost-sharing mechanisms, open-source solutions (e.g., C2PA), and the long-term benefits of transparency when developing policies and implementing practices. It's crucial to

find a balance between promoting transparency and ensuring that the costs are not prohibitive to innovation in the AI sector.

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

Encouraging compliance with transparency requirements in the context of AI and copyright necessitates a multi-pronged approach that combines incentives, education, and regulatory underpinning.

Incentives for Transparency

- **Recognition and Rewards:** Establishing programs that recognize and reward AI developers who demonstrate exemplary transparency practices can incentivize compliance. This could include awards, certifications, or preferential treatment in government procurement or funding programs.
- **Industry Standards and Certifications:** Encouraging the development of industry standards and certifications that signal transparency and ethical AI development can potentially lead to consumer trust and market advantages for compliant companies.
- **Reduced Legal Liability:** Offering reduced legal liability or safe harbor provisions for AI developers who proactively disclose training data sources and comply with transparency requirements can incentivize voluntary compliance.

Education and Awareness

- **Clear Guidelines and Best Practices:** Developing and disseminating clear guidelines and best practices for transparency in AI development, specifically addressing copyright considerations, is crucial. This could involve creating educational resources, workshops, and training programs for AI developers.
- **Public Awareness:** Raising public awareness about the importance of transparency in AI and the potential risks of copyright infringement empowers users to demand greater accountability from AI developers.
- **Collaboration and Knowledge Sharing:** Fostering collaboration and knowledge sharing between AI developers, researchers, and legal experts

promotes a better understanding of copyright issues and transparency best practices.

Regulatory Underpinning

- **Mandatory Disclosure:** Implementing regulations that mandate the disclosure of training data sources, particularly copyrighted works, with specific requirements for granularity and format, is essential. This is exemplified by the EU AI Act's emphasis on transparency and explainability, particularly for high-risk AI systems (Recital 133), and the UN's Resolution A/78/L.49 (see https://docs.un.org/en/A/78/L.49).
- **Enforcement Mechanisms:** Establishing clear enforcement mechanisms, such as audits, fines, or legal action, ensures compliance with transparency regulations and deters copyright infringement.
- **International Collaboration:** Promoting international collaboration and harmonization of transparency standards and regulations addresses the global nature of AI and copyright challenges. For example, the UN's International Telecommunication Union (ITU) is working towards industry-wide standards and protocols for content provenance and authenticity to ensure these tools work seamlessly across platforms and jurisdictions (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).

The Role of GDPR, CJEU, and C2PA

The GDPR's principles of data minimization, purpose limitation, and transparency provide a foundation for responsible AI development and data usage. The CJEU *C-634/21 SCHUFA Holding and Others (Scoring)* ruling's emphasis on transparency and explainability in automated decision-making systems reinforces the need for clear regulatory guidance in the context of AI and copyright. The Coalition for Content Provenance and Authenticity (C2PA)'s efforts to promote transparency and accountability in digital content provide valuable tools and standards for AI developers (see, e.g., https://c2pa.org/).

In conclusion, encouraging compliance with transparency requirements in AI development, especially regarding copyright, requires a multifaceted approach that

includes incentives, education, and regulatory underpinning. While voluntary measures can promote good practices, regulatory frameworks are crucial for ensuring accountability, protecting creators' rights, and fostering a responsible and transparent AI ecosystem.

28. What are your views on the EU's approach to transparency? Please give us your views.

The EU's approach to transparency in AI, as exemplified by the AI Act, specifically Article 53(1)(d) and Recital 107, which mandate the disclosure of copyrighted works used in training AI systems, is a positive step towards addressing copyright concerns in the age of generative AI. This approach aligns with the principles of transparency and explainability, allowing creators and users to understand how AI systems are trained and potentially identify instances of copyright infringement or unauthorized use of copyrighted material.

However, the effectiveness of this approach relies heavily on the specific implementation details. The level of granularity required in disclosing training data, as stipulated in Recital 107 of the AI Act, is crucial. It needs to strike a balance between providing meaningful information to assess copyright implications without compromising legitimate trade secrets or intellectual property. This means clearly identifying copyrighted works, highlighting significant contributions, and explaining data usage.

Accessibility and understandability of the disclosed information are also key, ensuring that complex technical details are presented in a clear and concise manner, potentially using standardized formats or visualizations (see https://c2pa.org/principles/). Robust enforcement mechanisms, as outlined in Articles 71 and 72 of the AI Act, are essential to ensure compliance with transparency requirements. This could involve audits, fines, or other penalties for non-compliance, particularly in cases of copyright infringement.

The EU's approach could benefit from global harmonization, as suggested by the ITU's report on the need for standards collaboration on AI and multimedia authenticity (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022). This would create a level playing field for AI developers and

promote international cooperation in addressing the challenges of AI transparency and copyright protection.

Challenges to the EU's Approach

Despite the positive intentions, several challenges could hinder the effectiveness of the EU's approach:

- **Defining the scope of disclosure:** Determining what constitutes a "trade secret" versus essential information for transparency can be tricky. Overly broad protection of trade secrets could undermine the disclosure requirements, while excessive disclosure might stifle innovation.
- **Specificity in disclosing copyrighted works:** The AI Act needs clear guidelines on the level of specificity required when disclosing copyrighted works, including identifying the specific works, their creators, and the extent of their use in training. Ambiguity could lead to inadequate disclosure or unnecessary burdens on AI developers.
- **Ensuring accessibility and understandability:** Presenting complex technical information about AI training and data usage in a clear and understandable manner for a diverse audience can be challenging.
- **Enforcement and compliance:** Monitoring and auditing compliance with transparency requirements, especially for a large number of AI developers, can be resource-intensive.
- **Cross-border challenges:** AI development and deployment often occur across borders, posing challenges for enforcing transparency requirements and addressing copyright infringement in different jurisdictions.
- **Global harmonization:** Achieving global harmonization is essential but challenging due to differing legal frameworks and priorities across countries.

Adapting to Evolving Technology

To address these concerns and navigate the complex landscape, the World Economic Forum advocates for strategic foresight. This proactive approach utilizes methods like scenario planning and horizon scanning to analyse future trends and

their impact. By engaging diverse stakeholders, governments and organizations can anticipate needs, adapt regulations, and address potential risks proactively. This agility is critical, as the rapid advancement of AI necessitates flexible and adaptable regulations. Policymakers must be responsive to new developments, considering diverse perspectives and exploring innovative approaches like regulatory sandboxes for controlled testing of new technologies (see https://www3.weforum.org/docs/WEF_Governance_in_the_Age_of_Generative_AI_2024.pdf). However, it is also worth noting that some newer AI style technologies, such as biological computers, may not come under the existing regulatory framework.

Addressing these challenges will require ongoing dialogue and collaboration between policymakers, AI developers, copyright holders, and other stakeholders. By proactively addressing these potential roadblocks, the EU can maximize the effectiveness of its transparency-focused approach and foster a responsible and innovative AI ecosystem that respects copyright while promoting technological advancement.

Wider clarification of copyright law:

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?

RIghtholders of artistic and literary works are the backbone of the UK's valuable, well- known (and emerging) creative and cultural works. The original literary and artistic works attract copyright protection for the life of the author plus 70 years and allow the creator to restrict acts in relation to their work around copying, distributing, and creating derivative works, for example. Moral rights around integrity of the work and attribution are also awarded. No additional hurtles, such as registration, are required, following our international obligations under TRIPS.

The exceptions and limitations in relation to copyright are designed to further the incentivisation and purpose of these limited economic and moral rights. The UK has taken a cautious approach to allowing use of copyrighted material without authors' permission by implementing a closed list of exceptions under 'fair dealing' and only gradually adding any further exceptions and limitation after careful research and consideration.

Do not create a blanket commercial *opt ou*t exception. This path maybe appear economically tempting to try to incentivize AI developers to establish and train their models in the UK, especially if other countries appear to offer this commercial training

exception. It cannot be overstated how this approach, in the long term, will be one of the most detrimental interventions possible to creators currently in businesses relying on copyright or for new companies and entrepreneurs looking to enter a creative market. A blanket commercial opt-out will affect and in some cases prohibit new entrants to the market

It is essential to consider the extraordinary time, human, political, and monetary resources at the disposal of the most active AI developers. A sweeping commercial opt out exception would decimate rights, protection, and human created artistic and literary development in the country, especially for smaller creators. An opt out system would be overly burdensome and unlikely to be a meaningful or understandable system for creators of varying degrees of sophistication around the law and varying experiences as to the possible returns when exercising these rights with their works, especially against a large or well resourced company. All rightsholders, not just companies, should have rights protected.

The Summary Assessment Annex 1 also cites that the Government's 2021 Consultation Report found respondents confirmed a reservation of rights is unduly burdensome. (*See* P5) There is no indication that this mechanism would be less burdensome or confer greater benefit now. The Summary Assessment also focusses on businesses. Many creators in the UK are not a 'business' and may not ever consider formalising economic exploitation of their work; their rights are no less important. Individual authors creating literary and artistic works would be deprived of copyright protections for their own works. Sacrificing artistic and literary creations for a purported short term gain is harmful.

The positive impact cited that 'Where such works are made available online they would need to be accompanied by a machine readable method to reserve rights, so that systems data mining significant numbers of works can easily identify works that can be lawfully mined' (p 4 Summary Assessment) would best be met by an opt in, not an opt out..

https://urheber.info/diskurs/ai-training-is-copyright-infringement#:~:text=The%20work%20of%20Prof.,of%20text%20and%20data%20mining

The current implementation of TDM was not designed to facilitate unfettered commercial exploitation of any works – copyrighted or not. The intention was to further science and research as training AI models does require material, but the wholesale copying and derivate works from copyrighted material without active intervention of rightsholders is contrary to the aims of balancing interests.

Although an 'EU-style opt out' is considered in the current proposal, the EU exception is far narrower, and most importantly, the architect now does confirm the legal intention was not to be used for such system. The lawmakers did not foresee or could not anticipate the shape and speed of ai modelling. (*see* https://www.theguardian.com/technology/2025/feb/19/eu-accused-of-leaving-devastating-copyright-loophole-in-ai-act)

The Summary Assessment of Options states that rightsholders would benefit with increased licensing activity in an opt out system would need significantly more research and support to warrant a change in the law with more consideration of the nuance and balance of rights and interest. For instance, evidence is provided as a wider benefit to society on productivity of works with cited study by NBER (FN 13 in the Summary Assessment of Options); this study provides evidence on call centre

productivity. Although certainly valuable research for its purpose, a pure economic analysis in relation to human creativity and legal protection will not capture the more complex and important factors when regulating this area.

Additionally, it is quite likely that much online creative material has already been subsumed into AI Models without permission. Thus, expanding a TDM now would not result in an uptake of licensing activity for the foreseeable future nor be likely to stimulate any new creative.

Creators might be deterred from exercising the 'opt out' option for several reasons more nuanced reasons, including hopes for future employment or collaboration, concerns about reputational harm, fear of missing out on competition or exposure, the belief that opting out may not be relevant, a preference for focusing on creation rather than technical or legal issues (which may be beyond their time and financial capabilities for questionable returns), a lack of understanding of the full implications of not opting out (especially as technology rapidly evolves), and discriminatory treatment for authors with limited internet access, education, or literacy.

<u>If A Commercial Opt Out Exception is Adopted</u>

If a commercial exception is nonetheless the chosen option, it is essential the exception to be opt in for rightsholders and to require AI developers to clearly record and disclose the sources of their training data as well as the opt in permission. This approach can be seen in works that are obviously in copyright but the rightsholder cannot be located (orphan works). An addition of a system similar to the Orphan Works Register could be utilised to ensure that creators and see how and where their copyrighted works have been used and then better understand whether they would like to opt out in the future or whether they are quite happy with attribution. The moral right of attribution here could be an interesting way to consider whether AI model training could more fairly balance the rights of creators so the works used to train have full integrity and recognition.

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?

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

The temporary copies exception should follow the existing legal precedent. Ultimately, training and AI model does not fall under a ‘temporary copy’ exception. The training process is not transient or incident and has extremely valuable economic significance in a commercial context, often as the basis of a derivative work. It could be worthwhile adding an advisory point of clarity. However, based on existing legislation and case law, an true representation of the technical process at issue should not even be considered to be within the realm of temporary copies, which were meant to essentially allow a technical process for copyrighted works to be made available digitally with no impact on the rights or work.

Question 27. If so, how could this be done in a way that does not undermine the intended purpose of this exception?

If the exception is clarified, the clarification should leave no ambiguity as to the intent of the law in the area of AI modelling.

Encouraging research and innovation

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

Research and innovation can be attracted and encouraged alongside protecting copyright. The TDM should be clarified to be for non-commercial purposes. The author of the TDM in the EU has confirmed that the TDM was misconstrued and was never meant to provide a broad commercial use exception. It was meant to balance interests and allow of personal use and progress. Any ambiguities around the intended application of the TDM could be clarified in light of the rapid recent technological developments that were not anticipated by the drafters, leading to unintended consequences.

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

This approach was attempted with the EUAIA on risk. Trying to legislate on the size or risk of a firm or model is probably not a useful approach as it is difficult to meaningfully anticipate the rapidly evolving issues overall. It may be important when developing regulation to consider the size of the firm in relation to ability to adhere to legal requirements and administer reports on transparency and training data. However, outside of the existing exceptions there is no need for additional exceptions related to copyright.

CGW Policy Option 0: No legal change, maintain current provisions

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.

No. It is incredibly difficult to see how s9(3) can ever be applicable to situations where works are produced using AI and which satisfy the originality requirement based on the judgment in THJ *Systems Ltd v Sheridan* [2023]. There is a substantive lack of clarity and lack of practical applicability with this provision. The approach to copyright requiring a human author in order to satisfy the intellectual creation requirement presents significant and substantial barriers to CGW benefitting from copyright protection. Given the way in which most current e.g., GenAI systems work, it is difficult to see that anything generated here could satisfy the originality requirement anyway.

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

No – see above.

CGW Policy Option 1: Reform current protection to clarify its scope

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

Yes. There is a need for clarity in respect of protections available for computer-generated works, even if this means removing the s9(3) element. The rise of AI in particular, and the rise of AI-generated works requires some form of intervention given the change in capabilities and uses since 2021.

Clarification should be offered as to on which basis computer-generated works are protected. Additional clarity is required in respect of how computer-generated works fit within the existing tests for originality (currently, they do not), or how they could fit within the entrepreneurial works provisions as a functional alternative. There is potential merit in considering computer-generated works to benefit from e.g., sui generis protection similar to that offered to databases under the CDPA.

33. Should other changes be made to the scope of computer-generated works protection?

No. We would not advocate for wide-ranging changes without a wider consideration of the scope of protection, and in particular, consideration being given to both the closed list of copyright categories, and the exemptions within the CDPA 1988.

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

CGW Policy Option 2: Remove specific protection for CGWs

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

No. This would not fit well with the Government's proposed AI Pathway, nor would it benefit creative industries nor authors.
To abandon protection entirely would shift copyright into a split scheme of protection under UK law, where there is an implicit quality threshold introduced which differs from the current difficulties around s9(3). It is likely that to remove the protection would lead to situations whereby there are bigger issues. This could also have unintended consequences in respect of infringement actions, licensing arrangements, and even situations of accidental inclusion for instance, all of which would have implications for human authors of other works, as well as for those human authors that rely on computer-generated works.

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

N/A

37. Would the removal of the current computer-generated works provision affect you or your organisation? Please provide quantitative information where possible.

N/A

Infringement and liability relating to AI-generated content

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

The current approach to liability for AI-generated outputs is still evolving, and its effectiveness in enforcing copyright is debatable. Several challenges exist:

- **Determining liability:** It's often unclear who holds the copyright to AI-generated outputs and who is responsible for infringement, as multiple actors (developers, providers, users) may be involved.

- **Adapting copyright law to AI:** Traditional copyright law focuses on human authorship, making it difficult to apply to AI-generated outputs. Questions arise about substantial similarity and derivative works in this context.
- **Balancing innovation and protection:** Overly strict liability could stifle AI innovation, while lax enforcement undermines creators' rights.
- **Jurisdictional variations:** Different countries have varying copyright laws and interpretations, creating challenges for international enforcement and harmonization.

For example, the US Copyright Office recently clarified that AI-generated output itself is not copyrightable, though a human author can receive copyright protection for creatively selecting or arranging AI-generated output (see https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf).

Current Approaches and Challenges

Current approaches often focus on the user's liability for infringing outputs. However, this may not be effective if the user is unaware of copyright implications or if the AI tool generates unexpected outputs. Targeting AI developers raises concerns about their control over outputs and potential chilling effects on innovation. Fair use/fair dealing exceptions and licensing agreements offer some flexibility, but determining fair use in AI training is complex, and licensing massive datasets can be expensive (see https://www3.weforum.org/docs/WEF_Governance_in_the_Age_of_Generative_AI_2024.pdf and https://www.law.berkeley.edu/wp-content/uploads/2024/04/Fair-learning.pdf).

In the US, *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, No. 1:20-CV-613-SB (D. Del. Feb. 11, 2025), addresses copyright and "transformative use" in AI training, emphasizing court scrutiny of training data under the "fair use" doctrine. The court's analysis focused on the fourth fair use factor (17 U.S.C. § 107) and concluded that the use was not transformative. This case, focused on non-generative AI, highlights the challenges of fair use in AI training and the need for AI developers to be cautious with training data. In the UK, the Copyright, Designs and

Patents Act 1988 (CDPA) is the core law, with key provisions regarding copyright ownership (e.g., sections 11 and 178) and exceptions for text and data mining (e.g., section 29A and CDPA provisions relating to fair dealing) being particularly relevant to AI.

The Path Forward

To improve the situation, policymakers need to:

- Provide clearer guidance on copyright ownership and liability in the context of AI-generated outputs.
- Explore technological solutions like Content Credentials and blockchain-based provenance tracking to identify the origin of AI-generated outputs and facilitate copyright enforcement (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).
- Develop industry standards for ethical AI development and data usage.
- Harmonize copyright laws and enforcement practices across jurisdictions (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022 and https://www3.weforum.org/docs/WEF_Governance_in_the_Age_of_Generative_AI_2024.pdf).

In conclusion, the current approach to liability in AI-generated outputs faces significant challenges in effectively enforcing copyright. A combination of clear legal frameworks, technological solutions, industry standards, and international collaboration is needed to navigate this complex landscape and ensure a balance between innovation and protection.

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

AI providers should adopt a multifaceted approach that prioritizes ethical data usage, transparency, and compliance with legal frameworks to avoid copyright-infringing outputs. This aligns with the UN's Resolution A/78/L.49 on Safe, Secure, and

Trustworthy AI, which underscores the need to protect intellectual property rights, particularly copyright (see https://docs.un.org/en/A/78/L.49).

Ethical Data Usage and Provenance Tracking

- **Meticulous Tracking:** AI providers should meticulously track the origin and usage of training data, ensuring that all copyrighted material is properly licensed and used with the creators' consent. This provenance tracking should document the source of the data, any modifications made, and the specific usage rights granted.
- **Content Credentials:** Integrating technologies like Adobe's Content Credentials can help verify the ethical sourcing of training data and ensure that creators are properly credited and compensated when their work is used. This metadata can be crucial in demonstrating compliance with copyright regulations (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).

Licensing and Alternative Data Sources

- **Proactive Licensing:** AI providers should proactively seek clear and comprehensive licensing agreements for copyrighted material used in training datasets. This includes understanding the scope of permitted use, attribution requirements, and any restrictions on commercialization.
- **Alternative Data Exploration:** To mitigate the challenges of licensing massive datasets, AI providers should explore alternative data sources, such as public domain works, openly licensed content, and synthetic data generation. This can help reduce reliance on copyrighted material while ensuring the availability of diverse and representative training data.

Transparency and Explainability

- **Transparent Data Usage:** AI providers should be transparent about the data used to train their models, including the source, licensing terms, and any potential copyright implications. This transparency can help build trust with

creators and users, fostering a more collaborative and responsible AI development environment (see https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/deepfake-detection_en and https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2023-11-16-techdispatch-22023-explainable-artificial-intelligence_en).

Engagement and Education

- **Policy and Legal Engagement:** AI providers should actively engage with policymakers and legal experts to stay informed about evolving copyright laws and regulations. This includes participating in discussions on fair use/fair dealing, licensing models, and other legal frameworks relevant to AI development (see https://www3.weforum.org/docs/WEF_Governance_in_the_Age_of_Generative_AI_2024.pdf).
- **Copyright Education:** Investing in copyright education and training for their teams is crucial, ensuring that all employees understand their obligations and contribute to a culture of copyright compliance.

As previously mentioned, the US case *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, No. 1:20-CV-613-SB (D. Del. Feb. 11, 2025), highlights the challenges in applying "fair use" to AI training. The court rejected an AI startup's defence for using copyrighted data, finding the use non-transformative because it mirrored the original work's purpose. This ruling, focused on non-generative AI, emphasizes the need for AI developers to be cautious with training data and consider licensing or uncopyrighted sources.

However, this may not be effective if the user is unaware of the copyright implications or if the AI tool generates outputs that are unexpectedly similar to existing works. This issue is highlighted in the Getty Images (US), Inc. v. Stability AI, Inc. cases. Targeting AI developers or providers raises questions about the extent of their control over the AI's output and the potential for chilling innovation. Existing secondary liability rules were discussed as a potential avenue for discussion in Getty.

That said, by adopting the measures above, AI providers can mitigate the risks of copyright infringement, foster innovation, and contribute to a more ethical and sustainable AI ecosystem. This proactive approach not only protects creators' rights but also promotes the development of trustworthy and reliable AI technologies that benefit society as a whole.

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

Yes, BILETA agrees that generative AI outputs should be labelled as AI-generated. This is crucial for transparency and accountability in the realm of copyright, allowing users to understand the origin of a work and assess its potential copyright implications. A proportionate approach to labelling should consider several factors:

Clarity, Standardization, and Accessibility

- **Clear and Concise Labels:** Labels should be clear, concise, and easily understandable, placed prominently to ensure visibility without disrupting the user experience. This aligns with the principle of accessibility of the European Convention on Human Rights (Article 8(2) and Article 10(2)) and European Court of Human Rights case law (see, e.g., *Kennedy v the United Kingdom* App no 26839/05 (2010) 52 EHRR 247).
- **Standardization:** Standardization, such as that promoted by the Coalition for Content Provenance and Authenticity (C2PA), is key, ensuring labels are consistent across platforms and jurisdictions to avoid confusion and ensure interoperability (see, e.g., https://c2pa.org/).
- **Flexibility and Adaptability:** The labelling approach should also be flexible, adapting to different content formats and AI technologies, accommodating the evolving nature of generative AI.
- **Accessibility for All Users:** Accessibility is vital, ensuring labels are accessible to all users, including those with disabilities or limited internet access (see https://c2pa.org/principles/).

The Need for Regulation

Regulation is likely required to ensure the consistent and effective labelling of AI-generated content, particularly in relation to copyright. This could involve:

- **Mandatory Labelling:** Laws or regulations requiring creators or distributors of AI-generated content to clearly label it as such, as exemplified by the EU AI Act (Article 50(4) and Recital 134).
- **Standardized Labels:** Regulatory bodies could establish specific standards for labels, including their format, content, and placement, especially regarding how they relate to copyright information.
- **Enforcement Mechanisms:** Regulations should include mechanisms for monitoring compliance and addressing violations, such as fines or other penalties for misrepresenting AI-generated works as original creations.
- **International Cooperation:** International cooperation is essential to harmonize labelling standards and ensure a coordinated global response to copyright challenges posed by AI-generated content.

The Role of International Frameworks

The UN's Resolution A/78/L.49 on Safe, Secure, and Trustworthy AI calls for robust, adaptable, accessible, and internationally interoperable tools for identifying AI-generated content (labelling and watermarking) (see https://docs.un.org/en/A/78/L.49), along with the EU AI Act's mandate for continuous refinement of these tools (EU AI Act Recital 133). These highlight the growing recognition of the importance of labelling in the context of copyright. While challenges remain, such as the potential for metadata alteration and the need for ongoing improvement of detection tools, a regulatory framework that promotes transparency and accountability in the use of generative AI is essential for navigating copyright concerns, protecting creators' rights, and fostering innovation in this rapidly evolving field.

Existing law in the CDPA 1988

Watermarking as currently construed under s.296ZG CDPA 1988 could play a role. The only addition would be to make it mandatory for all generative AI outputs. Removal of such a mark with knowledge would constitute an infringement of the provision. Currently the section is primarily aimed at the removal of copyright management information, but an AI label would invariably contain such information bringing it within the scope of the s.296ZG provision.

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

Governments can significantly support the development of AI content labelling tools and standards by fostering a coordinated, global approach. The rapid spread of deepfakes and synthetic media necessitates robust, standardized solutions, such as those built on the C2PA framework, which uses metadata, fingerprinting, and watermarking for content authentication (see, e.g., https://c2pa.org/).

Key Areas for Government Support

To effectively support these tools, governments should:

- **Invest in standardization:** Address C2PA's standardization challenges by promoting consistent label displays across platforms and funding research into simplified data interpretation, working with industry and international bodies to create and enforce standards.
- **Fund technical development:** Support the development of advanced detection methods, including cryptographic techniques, tamper detection, and blockchain-based provenance tracking, to overcome the limitations of current labelling systems.
- **Promote user education:** Combat user over-reliance on labels by funding public awareness campaigns that emphasize the importance of critical thinking, source verification, and contextual analysis.
- **Address scope limitations:** Encourage research into technologies that go beyond provenance, addressing manipulated meaning and contextual manipulation.

- **Protect privacy:** Develop and enforce regulations that balance the need for provenance data with user privacy, addressing concerns related to data collection and storage.
- **Support infrastructure:** Invest in the infrastructure needed to manage vast provenance records, ensuring efficient data storage and retrieval.
- **Combat data stripping:** Fund research into tamper-resistant labelling and watermarking techniques, and legislate against the intentional removal of provenance data (see https://reutersinstitute.politics.ox.ac.uk/news/spotting-deepfakes-year-elections-how-ai-detection-tools-work-and-where-they-fail).
- **Ensure accessibility:** Support the C2PA's commitment to accessibility by funding research into tools and interfaces that are usable by individuals with disabilities or limited internet access (see https://c2pa.org/principles/).
- **Foster adaptability:** Promote a dynamic ecosystem that includes decentralized platforms for information sharing, open-source intelligence, and AI-driven detection models that can adapt to evolving threats (see https://www3.weforum.org/docs/WEF_Governance_in_the_Age_of_Generative_AI_2024.pdf).
- **Drive international collaboration:** Support international interoperability by working with organizations like The UN's International Telecommunication Union (ITU) to harmonize legal frameworks, foster diplomatic collaboration, and establish clear data-sharing protocols, addressing geopolitical tensions and building trust between nations (see https://acrobat.adobe.com/id/urn:aaid:sc:EU:764a0bb2-52cc-4617-b8c3-690cf6f2d022).
- **Create adaptable legislation:** Following precedent set by cases such as *C-314/12 UPC Telekabel Wien GmbH v Constantin Film Verleih GmbH and Wega Filmproduktionsgesellschaft GmbH*, governments can create legislation that is adaptable and proportionate to the ever-changing risks associated with online content.
- **Embrace or avoid conflict with existing copyright management information laws** – laws which will have been introduced further to the WIPO Copyright Treat 1996. Countries with DRM provisions – which are generally

well known – will also have less well known copyright management information laws. However, those copyright management information laws will typically cover AI watermarks, as under s.296ZG CDPA 1988.

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

The EU's approach to AI output labelling, as exemplified by the AI Act (AIA), specifically Recital 133, rightly calls for flexibility to handle diverse content and evolving technologies. This is crucial for efficient compliance, especially when dealing with AI-generated content that might raise copyright concerns. Recital 133 emphasizes accurate tools like watermarks and metadata to trace content origin and authenticity, directly impacting the ability to determine if AI-generated works infringe on existing copyrights or fall under fair use exceptions like parody or satire. The C2PA's Content Credentials 'CR' icon, providing a detailed "digital X-ray" of content, aligns with this need for transparency, offering data that can be vital in copyright assessments – revealing creator info, creation date, tools used (including generative AI), and edits (see https://c2pa.org/post/contentcredentials/).

Clarifying Copyright Implications and the Role of XAI

However, the AIA must clarify how its transparency requirements address the nuanced area of copyright infringement, particularly in relation to deepfakes and AI-generated creative works. This is essential for distinguishing between legitimate parody/satire and outright infringement. Explainable AI (XAI) can play a critical role here (see https://www.edps.europa.eu/data-protection/our-work/publications/techdispatch/2023-11-16-techdispatch-22023-explainable-artificial-intelligence_en). By revealing how AI systems arrive at their conclusions, XAI can help determine if a work incorporates substantial elements of copyrighted material or if it constitutes a transformative use, such as parody. It can pinpoint specific features like altered melodies, character likenesses, or narrative structures, and provide human-readable explanations of the AI's reasoning. XAI techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) can identify the most influential features in the AI's output, aiding in copyright analysis (see https://arxiv.org/abs/2305.02012). This transparency, as demonstrated by companies like DuckDuckGoose (https://www.duckduckgoose.ai/)

and Reality Defender (https://www.realitydefender.com/blog/visual-deepfake-detection-explainability), is vital for navigating the complex intersection of AI-generated content, copyright, and creative expression, ensuring that legitimate creative endeavours like parody and satire are not stifled while protecting copyrighted works.

There should also be clarity over the relationship between AI output labelling and Article 7 of the 2001 EUCD, which provides protection for aforementioned copyright management information – which can (and likely does) overlap.

Digital replicas and other issues

Digital Replicas 16. The volume and quality of digital replicas generated by AI systems (sometimes called deepfakes) is increasing. This consultation seeks to gather evidence on the challenges posed by digital replicas. This responds to concerns around the emergence of deepfakes and AI-generated content that may replicate a person's voice, image, or personal likeness. It asks whether the current legal framework is sufficient to provide individuals with control over use of their likeness and whether further intervention is required.

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?

An opt out system with a broad commercial exception would provide individuals little control. This system would likely encompass personal photos shared on social media or on photos posted from work and events. Not all companies will respond quickly enough once the opt out system is in place nor consider it necessary. But the extent of the individual images online would make it overly burdensome to administer. Additionally, individuals have varying degrees of technical sophistication, understanding of the legal and social implications of opting out or not, confidence or capacity to negotiate licenses if asked, and also regular internet access. These factors can also be statistically linked to class and race, so this type of legal intervention could deepen inequalities.

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

AI and digital replicas pose several risks, including the spread of misinformation, personal and civil harms such as reputational damage, and economic damage. There is also the potential for criminal liability or regulation through existing instruments, such as the Online Safety Act or laws addressing criminal acts related to deepfake pornographic material, harmful media representations, and issues affecting young people, like pro-ana or radicalising content.

In relation to tortious remedies, we can offer remedies in the form of defamation, misuse of private information (privacy), malicious falsehood, passing off, or sometimes performers' rights. However, the laws were not written with these particular

types of issues interfacing with such sophisticated technology in mind, especially considering the ease and potential volume of harmful or unauthorised digital replicas.

If one of these works involving an unauthorised digital replica is copyrightable (or if the relevant digital replica aspects can be copyrighted), then the claimant might have a ground in moral rights of attribution or perhaps an extended ground in performers' rights.

In consideration of the stated goals of the consultation on protecting creators, artists, authors, and private individuals are not well served by being forced into a position of personal litigation with well resourced and financially lucrative large companies that may be developing the AI systems. This would be an area that would benefit from carefully considered enhanced legal regulation as the existing law will not fully protect individuals.

Other emerging issues

Impact on cultural heritage practices

The UK ratified the 2003 Convention for the safeguarding of intangible cultural heritage in June 2024. Intangible cultural heritage or 'living heritage' are living practices passed down through generations, reflecting the identity of a community. This might look like knowledge of crafts or textiles, traditional songs and dances, or other traditions around festivals or celebrations. From an economic viewpoint, the UK is an exceptionally rich variety of these heritages which provide reputational advantage and attract tourism. In accordance with Convention obligations, the UK has begun the process of consulting communities to determine the currently practiced living heritages, so our living heritages will take a forefront in the coming years. From a less direct or non-economic viewpoint, communities with strong identities and a strong sense of belonging are more productive, experience less crime, and report overall greater mental and physical well-being and health.

Living heritage is intangible practice. However, as much of our life is digital many of these living heritage practices will be recorded, explored, and even evolve organically through some form of digital medium. The positive outcomes and living heritage interacting with technology can be that endangered heritage might persist by reaching new practitioners and can evolve organically alongside communities and technology. Engagement with digital technology doubt does mean fixation in a tangible form of expression. If sufficiently original, the living heritage as a work may attract copyright and related exclusionary rights. Thus any digitisation should raise concerns around owning heritage that should belong to broader communities. Specifically in relation to generative AI, living heritage that is meant to evolve with the practicing community could be significantly influenced by AI adaptations and also could be impacted by an 'authorised heritage discourse' (*See* Smith, L. 'Uses of Heritage' (2006)) fixed as a source what the 'correct' heritage is rather than from generational knowledge. This is a more complex and subtle impact on human creative and artistic works but nevertheless foundational to considerations when regulating training and production of AI models in relation to culture and creativity.

The Role of Terms and Conditions

The role terms and conditions play here should not be underestimated in this conversation. Terms and conditions have become ubiquitous in using digital tools. Given the density, length, and ease of bypassing the contractual obligations before accessing a site or tool, users are not reading the terms before using games, tools, or websites. Users thus are agreeing to a variety terms

which will now likely include perpetual, non-exclusive, worldwide licences to the hosting company to train AI models or the company to sell this information to allow third party AI developers to train on any data or input from users. Existing tools that are very clearly using the user data to further train AI such as otter.ai, ChatGPT, or Stable Diffusion are in wide use amongst creative communities and in many workplaces. It is unlikely that most web users are fully aware of the implications around confidentiality and copyright—amongst other issues— that arise from using these products. Ultimately more consideration of how terms and conditions must be communicated to users especially around AI training is essential in this area. Due to the complexities of adhesive and opaque private contracting, regulation via copyright will likely be contracted around without intervention and terms and conditions and significantly improved awareness on the part of users who would also have to sacrifice the convenience and usefulness of the technology. Ideally the reasonable balance would be reached by allowing users greater power and insight when engaging with terms and conditions online to make fully informed choices about how the information is used. It is actually unlikely to be the case that all users with the decide to opt out which seems to be a concern around ensuring technological progress. Many people could be quite happy to have their information for certain used to train systems that they are receiving benefits from. However, this must be done with full transparency and full consent.

Awareness of Centralisation of Communication and Power

Technology companies working on AI have moved beyond contributing useful, creative or social tools. When regulating this area, it is vital to also acknowledge that digital systems and data are the backbone of social control and regulation right now; therefore, when looking at any type of exceptions or access, security, competitive monopolies, and political and financial corruption are all factors in decision-making that would typically not come into play with copyright law. The financial and political resources of many large AI development companies mean any bid to attract businesses to the UK must not be done without careful consideration of these risks. The reach of these companies is not just ‘social media’ or development of socially useful technology but also overarching control of communication which the UK has always taken a thoughtful and balanced approach to ensure objective and clear reporting of news and educational information. So it is essential that this persist and regulation is made with awareness that this technology and centralisation of power has a near future potential to have a greater impact even beyond creative works as can be seen on the global stage.

Environmental Issues

A related issue to attracting AI developers to the UK is environmental impact. Usage of generative AI tools as a significant environmental impact as currently designed. Beyond use if an AI developer also proposed to locate a datacentre in the UK to support their development a thorough, comprehensive environmental impact analysis needs to be produced and taken seriously even in the face of global pressures of an AI ‘race’ and the short-term financial and reputational advantage. Currently many data centres supporting AI training and development use vast amounts of electricity and water, especially in the US. Data centres are best placed in rural areas with ample space. However, this also presents a significant strain on resources for citizens some of whom are most vulnerable. Many communities with AI data centres are experiencing difficulties with regular electricity outages due to the strain on the system. Data centres must be cooled with water; some companies are buying communities drinking water in order to run the centres. This is not the only way to run these centres. For instance, DeepSeek in

China managed to run their centre by selective section rather than all at once, which can significantly reduce these demands. So, it is possible to run the centres in a more financially and environmentally efficient way. But there must be understanding into me and companies to prioritise this balance and efficiency.

---

[i] Margoni T., TDM and Generative AI: Lawful Access and Opt-outs (2024) (2) *Auteurs en Media*, pp. 175-188.

[ii] Mezei P., A Saviour or Dead End? Reservation of Rights in the Age of Generative AI (2024) 46(7) *European Intellectual Property Review*, pp. 461-469.

[iii] Margoni T., TDM and Generative AI: Lawful Access and Opt-outs (2024) (2) *Auteurs en Media*, pp. 175-188; Mezei P., A Saviour or Dead End? Reservation of Rights in the Age of Generative AI (2024) 46(7) *European Intellectual Property Review*, pp. 461-469.

[iv] Mendis D. White B and Hong D., *Copyright and Open Norms in Seven Jurisdictions: Benefits, Challenges and Recommendations* (January 2024) at https://www.knowledgerights21.org/reports/copyright-and-open-norms-in-seven-jurisdictions-benefits-challenges-policy-recommendations/ (102pp)

[v] Lewendowski A., How Copyright Law can Fix Artificial Intelligence's Implicit Bias Problem (2022) 93 *Washington Law Review,* 579.

[vi] Senftleben M., A Tax on Machines for the Purpose of Giving a Bounty to the Dethroned Human Author – Towards an AI Levy for the Substitution of Human Literary and Artistic Works at https://ssrn.com/abstract=4123309 *See also*, Senfleben M., Generative AI and Author Remuneration (2024) 54 *International Review of Intellectual Property and Competition Law*, pp. 1535-1560.

[vii] Mendis D. White B and Hong D., *Copyright and Open Norms in Seven Jurisdictions: Benefits, Challenges and Recommendations* (January 2024) at https://www.knowledgerights21.org/reports/copyright-and-open-norms-in-seven-jurisdictions-benefits-challenges-policy-recommendations/ (102pp)