Response to the UK Intellectual Property Office's Consultation on Copyright and Artificial Intelligence

24 February 2025

We, a group of organizations that represent hundreds of news outlets, support trustworthy media reaching millions of citizens in the UK and around the world, and are committed to democratic resilience of our information ecosystems and digital markets, submit this joint response to the consultation opened by the UK Intellectual Property Office (IPO) on the legal framework for AI and copyright. The Open Markets Institute and its Center for Journalism & Liberty (U.S.), the Independent Media Association (UK), Impress (UK), the Global Forum for Media Development (global), and the Danish Press Publications' Collective Management Organization (DPCMO) welcome the opportunity to inform a policy framework that will have widespread ramifications on the news and creative industries in the UK and around the world.

The work of our organizations underscores that this consultation is not just about copyright, it is about democracy and who controls the information ecosystem. The Center for Journalism & Liberty has pioneered efforts to understand the role of, and value that, journalism and other forms of high-quality data generate in AI systems and to ensure journalists and creators are fairly compensated for their vital work by authoring proprietary policy papers such as "What is the Value of Journalism to AI?" and publishing commentary in Brookings, *Tech Policy Press*, and *Washington Monthly*.

We call for the implementation of an opt-in protocol for AI training data collection to align more closely with copyright principles, reinforce the rights of creators, and compel technology companies to respect these rights. We believe that anything other than an opt-in, consent-based approach will fundamentally undermine the viability and competitiveness of the UK's creative industries in favor of foreign technology companies' operations and investment, further contributing to market imbalances and undermining domestic innovation. The implications of how the UK decides these questions will thus have broader repercussions on the resiliency and sovereignty of the information ecosystem and on democracy.

The major AI companies are set to collectively spend upwards of $1 trillion on AI development over the next five years,[1] profiting off the backs of creators and publishers whose works have been unlawfully scraped for training data. They have purposely exploited loopholes and grey areas in copyright law in an attempt to make unprecedented profits before the law catches up

[1] Erum Manzoor, "Comparing Major Companies' AI Spending in 2024 and the Challenge of Productionizing AI Solutions," *AIM Councils,* 6 November 2024, https://council.aimresearch.co/comparing-major-companies-ai-spending-in-2024-and-the-challenge-of-productionizing-ai-solutions/.

with them. They have silenced whistleblowers[2] who have understood the illegality of the companies' operations and tried to warn the public.[3]

The concept of copyright (also known in some languages as "author's rights"[4]) was established initially in order to award exclusive rights to the *author* of an original work to make copies, while third parties could only make copies with the permission of the authors. The widespread practice by the AI industry to use text and data mining (TDM) without the express permission or consent of authors has turned the principle of copyright on its head. If the UK IPO truly wants to uphold the principles and values of copyright, it should not allow this practice to continue unfettered and should return to a consent-based system.[5]

The UK's proposed approach clearly demonstrates a bias toward the AI industry, particularly dominant foreign incumbents, at the expense of its creative and information industries. There is no guarantee that, even if copyright policies were adjusted to be more "friendly" to the AI industry, AI companies would choose to perform their training activities in the UK. Meanwhile, domestic information producers and creators will suffer and be left vulnerable to the next inevitable technological disruption that will profit unfairly off of their intellectual and creative labor.

The UK's creative industries represent a crown jewel of its economy, producing some of the world's most celebrated cultural works across literature, art, music, and some of the world's most respected and trusted news sources. This informational and creative excellence requires careful stewardship to maintain its vitality.

While technology corporations speculate about potential transformative benefits of AI systems, the safety and effectiveness of generative AI relies heavily on access to high-quality training data.[6] As renowned computer scientist Peter Norvig is credited as saying, "More data beats better algorithms, but better data beats more data,"[7] showcasing the increasing importance of exceptional human-generated creative works. Therefore, establishing regulatory frameworks to ensure fair compensation for creators is not just about protecting their rights – it is also about preserving the very source of innovation that AI companies depend upon for advancement and ensuring creators are still incentivized to create.

---

[2] Alys Davies, "OpenAI whistleblower found dead in San Francisco apartment," *BBC,* 14 December 2024, https://www.bbc.com/news/articles/cd0el3r2nlko.
[3] Suchir Balaji, "When does generative AI qualify for fair use?", 23 October 2024, https://suchir.net/fair_use.html.
[4] World Intellectual Property Organization, "Understanding Copyright and Related Rights," 2016, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.
[5] Courtney Radsch, "The case for consent in the AI data gold rush," *Brookings*, January 16, 2025, https://www.brookings.edu/articles/the-case-for-consent-in-the-ai-data-gold-rush/.
[6] Courtney Radsch, "AI Needs Us More Than We Need It," *Washington Monthly*, October 29, 2024, http://washingtonmonthly.com/2024/10/29/ai-needs-us-more-than-we-need-it/.
[7] Martha Amram, "How Generative AI Makes Sustainability Data More Valuable," *Forbes,* November 22, 2023, https://www.forbes.com/councils/forbestechcouncil/2023/11/22/how-generative-ai-makes-sustainability-data-more-valuable/.

Note: The submission below addresses only those consultation questions that fall within our core expertise in competition, market structure, and the economics of journalism and creative industries.

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

This opt-out approach flips copyright on its head and puts the burden on right holders – including independent content creators and under-resourced news organizations. It unfairly prioritizes Access over Control and Transparency, which together form three supposedly equally important objectives.

Option 3 is not a workable compromise between copyright holders and tech companies, as presented by the IPO. If this approach is adopted, the government risks falling into the same trap as the EU did with its 2019 Copyright Directive. The EU standard allows AI companies to use copyrighted content unless explicitly forbidden, yet there remain hundreds of AI crawlers[8] and it is virtually impossible and/or prohibitively expensive to keep up with the constantly evolving list of bots. Furthermore, many bots do not identify themselves or bypass and ignore instructions (link to MISO research when we get it).

2) Which option do you prefer and why?

The IPO should adopt Option 1 and require an explicit opt-in consent approach.[9] By doing so, the IPO can reinforce the principle that content creators have ultimate control over usage of their work in the context of data mining and AI training. It would also incentivize technology companies to develop innovative tools and methods to respect the rights of creators and publishers.

This option not only respects the original principles of copyright. It also takes into account the practical considerations of resources, knowledge, and time. The tech companies that have dominated the creation of the leading proprietary and "open source" foundation models have unprecedented resources, and they have top legal teams to navigate legal frameworks.

We consider the rights of creators and the balance needed to ensure a balanced market in which data providers receive fair compensation, just as the utilities, chip manufacturers, cloud providers, and other aspects of the AI tech stack do, to be the most important issues. The UK is uniquely positioned to play a leadership role in shaping how responsible and safe AI is developed while growing its flourishing creative economy. Creative industries will perish if companies are allowed effectively free rein to continue their unfettered use of data, including to develop market-replacing products.

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

---

[8] See for example Dark Visitors, "Agents," accessed February 21, 2025, https://darkvisitors.com/agents.
[9] Courtney Radsch, "The case for consent in the AI data gold rush," *Brookings*, January 16, 2025, https://www.brookings.edu/articles/the-case-for-consent-in-the-ai-data-gold-rush/.

We do not support an exception for commercial purposes. Commercial enterprises should be required to pay the costs it takes to develop and run a business. We do not believe that AI businesses should get out of paying for chips or should get energy, land, or staff without paying for it. Why should we encourage the development of a sector that is predicted to disrupt the entire economy that is not required to fully account for the costs of the products and services it develops? Furthermore, commercial AI companies, in particular the dominant technology corporations that overwhelmingly control AI technologies and AI unicorns, are among the most valuable companies globally and have shown their ability to attract funding despite concerns about data provenance, model collapse, and labor impacts. There is a symbiosis between human content creators and the AI industry that should be cultivated through balanced, consent-focused frameworks.

4) 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 suggest Option 1, or another similar opt-in approach. Again, we reiterate that only an opt-in approach can respect the original principles of copyright and the reality that right holders will not all understand the process to reserve their rights. Furthermore, there is no guarantee that technology companies with a history of skirting the law will respect opt-out reservations. Many already blatantly break the law, such as Meta when it knowingly used pirated books to train AI models[10] or Perplexity when it ignored do-not-crawl instructions.

The burden should be placed on the AI industry, which has more resources, staff, and time, and is equipped to navigate various regulatory frameworks already. The AI industry is currently valued at £72 billion[11] and is attracting £200 million a day in private investment.[12] At this rate, the AI industry will soon outpace the UK creative economy, valued at £124 billion.[13]

The UK has a thriving tech sector, valued at $1 trillion, yet startups reportedly aspire to get bought up by Big Tech rather than compete. Its vibrant cultural industries and public service media are among the best in the world, yet they have become mere inputs for Big Tech's AI products. AI systems trained on vast quantities of copyrighted material acquired without permission or compensation could render entire sectors of the economy obsolete.

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

The entity that ignores a reservation should be served a fine as well as the reimbursement of all profits from infringing activities and deletion of all unauthorized copies of copyrighted material

---

[10] Dan Milmo, "Zuckerberg approved Meta's use of 'pirated' books to train AI models, authors claim," *The Guardian,* January 10, 2025, https://www.theguardian.com/technology/2025/jan/10/mark-zuckerberg-meta-books-ai-models-sarah-silverman.
[11] Sofia Villegas, "UK AI sector most valuable in Europe, new report reveals," *Holyrood*, June 11, 2024, https://www.holyrood.com/news/view,uk-ai-sector-most-valuable-in-europe-new-report-reveals.
[12] UK Government, "UK AI sector attracts £200 million a day in private investment since July," press release, January 15, 2025, https://www.gov.uk/government/news/uk-ai-sector-attracts-200-million-a-day-in-private-investment-since-july.
[13] UK Parliament House of Lords, "Creative industries: Growth, jobs and productivity," January 30, 2025, https://lordslibrary.parliament.uk/creative-industries-growth-jobs-and-productivity/.

from systems and databases, including from datasets used for AI model training. Together, these remedies would both eliminate unauthorized content and remove financial incentives for infringement.

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.

While we believe that only an opt-in protocol will best achieve the goal of preserving the creative industries, if a rights reservation system must be adopted, rights should be reserved in machine-readable format (as in the EU Copyright Directive). However, with the Directive, there are no set standards or protocols for right holders to reserve their rights, and it is unclear how rights reservations expressed through a variety of methods are respected by AI companies.

The IPO should publicly identify data sources, protocols and standards for the expression of rights reservations that are freely available and whose functionality is publicly documented.[14] By doing so, the IPO will provide more clarity and guidance to right holders and make it easier for AI companies to comply. By providing a freely available (e.g. open source) standard for rights reservation, the cost of introducing and complying with the standard should be minimal. Development of technical standards should ideally be done through, or harmonized with, international standards-setting efforts at the Internet Engineering Task Force.

Furthermore, we highlight that implementing an opt-in protocol would be significantly more cost-effective than an opt-out protocol. Although we recommend that an opt-in protocol also use machine-readable formats compliant with a standard set by the IPO, the financial and administrative burden would be reduced since right holders would only need to apply the standard to works they specifically choose for AI training.

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

As stated in response to the question above, there is a need for greater standardization of protocols for rights reservation and for opt-in protocols. Currently, there is no standardization, which creates uncertainty for both creators and AI companies.

Standardizing rights reservation protocols benefits both right holders and AI companies. Focusing on a small number of standardized identifiers will provide more legal certainty and streamline opt-out processes for right holders. For AI companies, this standardization will reduce implementation complexity and cost.[15]

Machine-readable rights reservations should be freely available, contain functionality documentation available online, and be user-friendly for right holders. They must be able to

---

[14] Paul Keller and Zuzanna Warso, "Defining best practices for opting out of ML training," *Open Future,* September 28, 2023, https://openfuture.eu/publication/defining-best-practices-for-opting-out-of-ml-training/.
[15] Paul Keller, "Considerations for Opt-Out Compliance Policies by AI Model Developers," *Open Future,* May 16, 2024, https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf.

differentiate between various uses, such as AI training and indexing for search engines. For creators that allow the use of works subject to remuneration – such as through remuneration of specific works, licensing agreements, and other collective agreements – these machine-readable rights reservations would ideally be combinable with remuneration frameworks. We would draw your attention to the IETF process underway to develop an AI preferences protocol.

10) How can compliance with standards be encouraged?

To encourage compliance by AI companies, especially startups, the IPO and relevant authorities should launch technical assistance programs to spread awareness and knowledge of measures to be taken. Through trainings, webinars, communications campaigns, and office hours, staff at these agencies should inform representatives at these companies about the processes to comply with standards, the consequences if measures are not taken, and how to implement processes and tools for complying with rights reservation mechanisms.

To encourage compliance by right holders, communications campaigns should be launched to inform right holders of their rights and show them how to implement machine-readable rights reservation systems. Technical assistance programs should be launched in collaboration with trusted intermediaries — publishers' associations, collective management societies, industry bodies, and creative sector organizations — to ensure the guidance reaches creators across different sectors and scales.

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

The UK government plays a vital role in supporting the development and enforcement of the standard. Experts agree that rights clearance solutions should be the result of collaboration between the tech industry and the creative industries.[16]

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

Licensing gives power to creators and operates within a free market system, but it is not a replacement for public policy frameworks that provide legal clarity.

However, not enough creators will have sufficient bargaining power to strike licensing deals or be able to work with third-party representatives. Licensing deals tend to favor bigger players who not only have more bargaining power but also the capacity to engage in dealmaking (e.g. staff, expertise, technical capacity, etc.). Furthermore, smaller AI developers, research entities and non-profit developers are unlikely to have the resources to engage in direct negotiations with individual content providers, making it imperative that a clear licensing framework be established that negates the need for individual negotiations with each rights holder.

Furthermore, we are concerned that in the future the most well-resourced tech companies could strike exclusive content licensing deals. As publicly available data becomes scarce and the

[16] Martin Senftleben, "Generative AI and Author Remuneration," *IIC - International Review of Intellectual Property and Competition Law*, 7 November 2023, https://link.springer.com/article/10.1007/s40319-023-01399-4.

demand for human-generated data skyrockets, companies could seek out exclusive deals with publishers or content creators, giving them an unfair advantage over rivals.

14) Should measures be introduced to support good licensing practice?

Measures should be introduced to support good licensing practices, especially measures that preserve the balance of power between well-resourced companies and smaller actors, including startups and SMEs. Measures should require that licensing deal terms be transparent in order to give creators, especially smaller actors, the opportunity to strike similar deals.

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 encourage collective licensing and data aggregation services by establishing frameworks to ensure that creators have the information they need to effectively determine the value of their data and establish terms that are fair so that small/independent creators as well as a wide range of AI companies and organizations can benefit from these services. It can set regulatory frameworks for collective management organizations to ensure fair representation and transparent fee structures; provide technical support to establish digital licensing platforms; and create standards to enable efficient licensing.

The government should impose transparency requirements on AI models, RAGs, and other data uses that ensure content creators and data providers can determine when their data/content is used and ascertain appropriate valuation.

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

The data shortage is a well-known problem, and the largest providers of GPAI models have found workarounds such as unfettered web scraping, licensing partnerships and using proprietary and synthetic data.[17] However, these solutions are often opaque and require vast amounts of capital – and this effectively creates a "data moat" that privileges the biggest companies and puts up immense barriers to entry, harming competition in AI.

We support transparency into the datasets and sources used in training, testing, validation, and retrieval-augmented generation (RAG). This transparency will allow authorities to better understand where concentration of power in data exists and take appropriate action.

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

In line with the EU AI Act General-Purpose AI Code of Practice, Measure 1.1 in line with Annex XI §1 2.(c) and Annex XII 2.(c), we recommend the UK government require the following information to be submitted to the relevant authorities regarding data inputs used for training, testing, and validation of generative models:

[17] Courtney Radsch, "Dismantling AI Data Monopolies Before it's Too Late," *Tech Policy Press*, October 9, 2024, https://www.techpolicy.press/dismantling-ai-data-monopolies-before-its-too-late/.

- A comprehensive list of data acquisition methods including web crawling, private licensed data, third-party data, data annotation/creation, synthetic data, user data, publicly available data, and other collection means
- Detailed time periods for data collection across each acquisition method, with notation for any ongoing collection processes
- Data processing methodology used to transform acquired data into model training data, including preprocessing steps and validation procedures
- Web crawler specifications including user-agent strings, collection periods, operating organizations, and robots.txt compliance protocols
- Organizations managing human data creation/annotation, including location details and workforce numbers involved in the training process
- Previously acquired data usage methodology, including rights acquisition documentation and relevant product/service data sources
- Synthetic data generation methods and systems, including names of AI models used in the generation process
- Content filtering and safety protocols addressing copyrighted materials in the training data
- Legal compliance measures implemented during data acquisition and processing, including handling of sensitive information and bias mitigation strategies
- Dataset specifications including total size per modality and distribution across different acquisition methods and sources for training, testing, and validation sets

In addition, we recommend the following additions:

- Transparency on all of the above with respect to data used for RAG
- For data obtained through partnerships or licensing deals: Nature of the partnership (e.g., exchange for services, purchasing of data)
- Whether the data includes personally identifiable information and measures taken to anonymize data

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

In line with the above recommendations, we urge the UK government to require disclosure of all use of web crawlers and specifications of the web crawlers, including user-agent strings, collection periods, operating organizations, and robots.txt compliance protocols. We see this as a vital first step toward increasing transparency around the processes used by AI companies to mine and profit off creators' data. These should be reported to and available in a central machine-readable repository.

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

Approaches to transparency should take into account (1) the capacity of organizations and companies of differing size; and (2) the importance of maintaining a balance betweentransparency with the need to protect information security and trade secrets.

First, transparency measures should not disproportionately burden SMEs and startups. Disproportionately impacting SMEs and startups is likely to worsen market concentration and privilege providers that have the resources and time to comply with measures. Measures should largely target the largest providers of general-purpose systems. SMEs and startups should still be required to take transparency measures but at a lighter level, while ensuring that they respect copyright.

Second, to protect trade secrets, the information regarding transparency measures should only be shared in full with the appropriate authorities, including the UK IPO, ICO, and CMA, who can take action if they determine rights have been violated or abuses of power have occurred.

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

Compliance with transparency requirements cannot be encouraged; they can only be mandated and enforced. As mentioned, AI companies have long profited from the illegal scraping of data on the Internet, including millions of copyrighted works. They will not be incentivized to stop this profiteering through soft law approaches alone. Only enforcement of hard law can prevent tech companies from causing further harm and perpetuating a business model that is inherently extractive and anticompetitive.

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

The EU has taken a laudable step towards increasing the transparency requirements from technology companies with its new digital regulations, including the AI Act, Digital Markets Act (DMA), and Digital Services Act (DSA). They have increased transparency regarding their business practices and towards the public, including users, civil society, researchers, and competing companies. For example:

- The DSA requires that online platforms publish reports on their content moderation practices; and imposes further transparency requirements on very large online platforms and search engines (VLOPs and VLOSEs).[18]
- The DSA requires VLOPs and VLOSEs to allow vetted researchers to access platform data when the research contributes to the detection, identification, and understanding of systemic risks in the EU.[19]
- The DMA requires gatekeepers to provide access to performance-measuring tools and advertising data to advertisers and publishers.[20]
- The AI Act requires developers of systems that interact with humans to inform people that they are interacting with AI. It also requires deployers of AI systems that publish text that informs the public to disclose if the text has been artificially generated or manipulated.[21]

[18] DSA Articles 15 & 42.
[19] DSA Article 40(4).
[20] DMA Article 6(8).
[21] EU AI Act Article 50.

These regulations request information vital to determining copyright violations, preserving human rights, and ensuring open access to information that can combat dominant technology corporations' monopoly power. The increased transparency has made corporations more accountable and led users to make more informed decisions.

In addition, the limitations on access only for researchers, as opposed to journalists and civil society, are insufficient. Data access and transparency rights should be ensured for a wider array of interested parties.

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?

Encouraging AI developers to train their models in the UK is beneficial both from a national economic perspective and from an AI safety perspective due to the strength of the AI Security Institute and its partnerships with AI companies. However, the UK should not encourage this through loosening copyright laws and prioritizing AI companies over right holders. While this approach might attract more AI model training to the UK, it risks companies exploiting UK infrastructure and creative works for training while maintaining their primary operations elsewhere, leaving UK rights holders harmed with little lasting economic benefit to offset their losses. Steps the government can take to encourage UK-based AI training should focus on creating genuine long-term business presence, not just opportunistic model training.

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?

Rather than weakening UK protections to match more permissive jurisdictions, the UK should take a leadership role in developing international standards that protect rights holders, such as by adopting an opt-in, consent-based approach. Furthermore, even in seemingly more permissive jurisdictions, such as Japan, there is a reassessment underway given the deleterious impact on rights holders, particularly the news industry and creative industries. The UK should continue to work through international forums (G7, G20, etc.) to align approaches on copyright that prevent regulatory arbitrage while ensuring fair compensation for creators.

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

The temporary copies exception fundamentally cannot apply to AI training copies because of the core requirement in the Copyright Act that such copies "have no independent economic significance."[22] AI training copies are, in fact, economically critical because they form the foundational basis for AI models and therefore the AI companies' entire business models. Furthermore, unlike truly temporary copies (like browser caches), training data copies create lasting economic value through their role in model development.

The TDM exception also does not apply for generative AI training because TDM

[22] Copyright, Designs and Patents Act 1988, Schedule 2.

is about processing and analyzing semantic information to generate new knowledge, while generative AI aims to produce data similar to the training data.[23]

The temporary copies exception as it stands is currently exploited and misconstrued by AI companies. Clarification is needed to explicitly state that AI training activities are not covered under the temporary copies exception.

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

Clarification should explicitly state that AI training copies fall outside this exception precisely because they have substantial independent economic significance. It should also reinforce that the exception only applies to genuinely transient technical processes (like browser caches) for non-commercial purposes. Furthermore, it should make clear that attempts to characterize AI training copies as "temporary" would undermine both the letter and spirit of the exception.

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

There are two problems with this approach. First, the lines between non-commercial and commercial research remain blurred and difficult to enforce. For example, OpenAI started as a non-commercial organization and now has transitioned to a commercial enterprise – other companies may look to exploit this loophole to engage in TDM with the intention to commercialize later on. In addition, technology corporations often fund research, create NGOs, and establish academic positions to perform research, and it is often untransparent which company is behind a specific study.[24]

Second, the exception may not apply to journalists. Although they often work for commercial enterprises, journalists have an important role in providing information for the public interest – and often investigative research requires access to quality datasets and data-mining tools.

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

Both size of an AI firm and purpose of an AI model should be considered. As stated in the previous response, there should be a specific carve-out for data mining for the purpose of furthering information in the public interest. In addition, copyright rules should consider the size of an AI firm and should not disproportionately burden SMEs and startups so as to not worsen market concentration in the sector.

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

---

[23] Dornis, Tim W., "The Training of Generative AI Is Not Text and Data Mining," *European Intellectual Property Review*, forthcoming February 2025, https://claude.ai/chat/bfe7d957-2137-4cee-8bc9-899be565d108.
[24] Marietje Schaake, *The Tech Coup,* Princeton University Press, 2024.

Liability should apply to both generative AI providers and users that knowingly use plagiarized or copyright infringing outputs in a non-personal way, particularly in instances where copyright is infringed for *commercial* purposes. It is the responsibility of providers to ensure that their products are compliant with laws and standards, including copyright laws. Providers should be liable for outputs in cases where they did not receive permission or licensing for the inputs. They should also be responsible for ensuring that their models cannot reproduce direct copies of copyrighted material and that adequate measures are taken to avoid "jailbreaking" the models to remove safeguards. If their product does not meet these standards, the provider should be held fully liable for breaking copyright rule.

Liability could be placed on the user if they deliberately tamper with the model – especially open-source or open-weight models – and remove any safeguards set up by the provider or if they knowingly reproduce copyright protected material for uses not protected under fair dealing exceptions or for commercial purposes. In these cases, enforcers should set clear rules on the obligations of generative AI providers to prevent this from happening and the specific cases in which users could be held accountable.

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?

The current approaches to transparency and TDM are insufficient to provide individuals with control over their likeness, awareness into how it is being used, and visibility into the outputs. Current approaches focus broadly on copyrighted works but need specific provisions for personal likenesses, which are important not only from an economic perspective but also from a privacy perspective. Non-consensual, sexually explicit deepfakes have ruined lives, careers, relationships, and families. Impersonation and other generative AI creations may also detract from the ability of performers, public figures, and other people to exercise their rights of publicity in terms of "name, image, and likeness."

While the UK government is commendably implementing laws to make creating deepfakes a criminal offense,[25] a regulatory approach to control the inputs into these deepfakes is also important to prevent harms. A particularly robust opt-in approach should be adopted for any instance where personal data, especially image and voice, could be used for AI training. Technical standards for identifying and respecting these opt-ins should specifically address personal likenesses and personal data.

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

It is not clear how the current or proposed legal framework would treat AI products that interact at the point of inference, such as retrieval-augmented generation (RAG) systems. RAG systems should, at minimum, follow the same copyright rules as AI training, but arguably require even stricter oversight because they actively and continuously access copyrighted materials; directly

[25] Ministry of Justice, "Government crackdown on explicit deepfakes," press release, January 7, 2025, https://www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes.

monetize the use of copyrighted content and divert traffic and revenue away from publishing platforms; and present a higher risk of misinformation or hallucination, especially when RAG is used to understand news and current events.

In order to mitigate these risks, a framework that requires explicit licensing for RAG systems and properly attributes and compensates creators at the time of retrieval. Creators and publishers should also be able to access a mechanism to opt in to having their content used for specific purposes, including RAG.

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

The use of synthetic data to train AI models privileges the largest technology corporations, further entrenching data monopolies.[26] As the AI Now Institute observed, corporations with the "widest and deepest" data advantages will be able to "embed themselves as core infrastructure" and become indispensable and irreplaceable in the AI ecosystem.[27] The creation and use of synthetic data in itself constitutes a moat, since the development of high-quality synthetic data is technically complicated and favors established companies with many resources and top talent, especially those that have already strip-mined the internet without adhering to copyright and privacy rights.

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

The next frontier in AI technology will be AI agents. All existing and proposed regulatory frameworks should consider their adaptability to agents, especially fully autonomous ones. AI agents will present new challenges: unpredictable behavior, little to no human oversight, lack of transparency and accountability, and new questions over liability.[28] While there will soon be a need for an overarching framework to regulate AI agents and respond to these challenges, any regulatory frameworks will have to consider who will be liable when agents violate laws such as copyright. Copyright laws should still apply to activities performed by agents, and the principle of assigning liability to AI companies over users should be applied as much as possible.

###

Signatory Organizations:

- The **Open Markets Institute** and its **Center for Journalism & Liberty (CJL)**, based in Washington, D.C. and Brussels, Belgium, is a non-profit organization dedicated to

[26] Courtney Radsch, "Dismantling AI Data Monopolies Before it's Too Late," *Tech Policy Press*, October 9, 2024, https://www.techpolicy.press/dismantling-ai-data-monopolies-before-its-too-late/.
[27] Amba Kak and Sarah Myers West, "AI Now 2023 Landscape: Confronting Tech Power", AI Now Institute, April 11, 2023, https://ainowinstitute.org/2023-landscape.
[28] Tara S. Emory, Maura R. Grossman, "The Next Generation of AI: Here Come the Agents!", *National Law Review,* December 30, 2024, https://natlawreview.com/article/next-generation-ai-here-come-agents.

promoting fair and competitive markets. Its mission is to safeguard our political economy from concentrations of private power that undermine fair competition and threaten liberty, democracy, and prosperity.

- **Impress** is the UK's only recognized independent self-regulatory body for the press that regulates more than 200 news brands which reach more than 20 million monthly readers, ensuring they have access to trustworthy and responsible journalism.

- The **Independent Media Association** is an association of over 90 UK independent media publishers with a collective reach of over 12.5 million. It is the most innovative and growing part of the UK media sector and is championing a fair playing field so that innovation and growth can continue.

- The **Danish Press Publications' Collective Management Organization (DPCMO)** is a collective management organization representing about 99 percent of the Danish media and press publishers.

- The **Global Forum for Media Development (GFMD)** is the largest global community for media development, media freedom, and journalism support with more than 200 member organizations from more than 70 countries.