Kairoi Ltd
71-75 Shelton Street
Covent Garden
London
WC2H 9JQ

Intellectual Property Office (IPO)
Department for Science, Innovation and Technology (DSIT)
Department for Culture, Media and Sport (DCMS)

Sent via online form

20th January 2025

RE: Consultation on Copyright and Artificial Intelligence

Kairoi Ltd ("we" hereafter) warmly welcomes the government's invitation to provide input for the consultation on copyright and artificial intelligence ("the consultation" hereafter).[1] The consultation seeks views on how the government can ensure that the UK's legal framework for AI and copyright supports the UK creative industries and AI sector together. The government's objectives will entail a complex balancing of different interests; namely, enhancing right holders' control over whether or not their works are used to train AI models, and ensuring AI developers have access to high-quality material to train leading AI models in the UK, whilst boosting trust and transparency between sectors.

The feedback below follows the structure of the online form provided by the government, focusing on a subset of questions. For the purpose of this response, we have made every effort to account for the government's AI Opportunities Action Plan.[2]

Copyright and Artificial Intelligence

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

1. No. The first objective —supporting right holders' control of their content and ability to be remunerated for its use— is heavily undermined by this option. Option 3 places the onus on the rights holder to opt out from their creative outputs being used by third parties without their permission. Option 3 effectively furthers the issue described in paragraph 6: "creative and media organisations are concerned that their works are used to train AI without their permission." In the same way that data protection regulations require action to signify the

---

[1] IPO, DSIT & DCMS (2024) Open consultation: Copyright and artificial intelligence, *gov.uk*, online [accessed 01 January 2025]

[2] DSIT (2025) AI Opportunities Action Plan, online [accessed 13 January 2025]

granting of consent, we would strongly advise that permission be actively given rather than granted by default.

2. In the meantime, option 3's impact on objective 3 —promoting greater trust and transparency between the sectors— is ambiguous. No details are presented about the "robust measures to ensure developers are transparent about the works their models are trained on, and right holders either individually or collectively can easily reserve their rights." On the one hand, transparency measures could —as noted— be enabled by "dispute resolution mechanisms and oversight." However, it is immediately claimed that "this may be achieved in different ways in alignment with or even outside the existing copyright framework." Thus, it is unclear what a copyright framework within the scope of this consultation could achieve.

3. On the other hand, the measures for creatives to reserve their rights reinforce the interpretation that the onus is on them to protect their work. Conversely, the AI industry seems to be disproportionately shielded by the consultation. Consider that paragraphs 57 and 95 both speak of the need to mitigate the risks that "AI developers, […] small firms and new entrants in particular," with little mention of the detrimental effects to creatives — whose employment may already be precarious.[3]

Question: Which option do you prefer and why?

4. Option 1: Strengthen copyright requiring licensing in all cases. This option aligns most closely with a basic tenet we promote at Kairoi: we must develop and adopt AI in ways that do not infringe on people's rights. What option 1 protects is the rights of creatives and the spirit of copyright law. As stated in paragraph 38 of the consultation, "the purpose of copyright is to protect creative expression and thereby encourage creativity and investment in the creative industries."

5. The strongest argument against option 1 may be a presumably negative effect on the competitiveness of the UK's AI industry. Through previous consultations, Kairoi has already challenged the UK's "pro-innovation" approach to AI regulations.[4] Given recent developments in the field of AI, it is worth emphasising that innovation and regulation are not at odds.

6. "Small language models" —as opposed to "large language models" (LLMs)— are on the rise.[5] These are models that are more targeted and effective in their use, built on fewer data and with fewer parameters, and thus consuming less energy. Small language models are a great innovation that can result from more constraining legal environments; the data AI developers shouldn't be everything-by-default but what they have permission to access. In a world where the government takes forward option 1, AI developments coming out of the

---

[3] Jeary, L. & Brawley, S. (2024) Cultural and creative industries: resources, funding, support and future, *UK Parliament*, online [accessed 01 January 2025]

[4] Kairoi Ltd (2023) RE: Consultation on policy proposals for UK's pro-innovation approach to AI Regulation, *Github*, online [accessed 01 January 2025]

[5] DeBiase, D. (2024) Why small language models are the next big thing in AI, *Forbes*, online [accessed 01 December 2025]

UK can be some of the most innovative: we would create powerful models that are effective and respect people's rights to create and own artistic content.

Exception with Rights Reservation

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

7. No. Kairoi has previously commented on introducing exceptions to text and data mining (TDM) regulations.[6] In the House of Lords' report following that same inquiry, we find a key statement that aligns with what we stand for at Kairoi: "The point of copyright is to reward creators for their efforts, prevent others from using works without permission, and incentivise innovation."[7] Although section C mentions remuneration being offered in exchange for licences, it continues to burden the creative industry with the need to negotiate such licences. As mentioned above, we maintain that this disproportionately impacts creatives whilst supporting AI developers.

Question: What other approach do you propose and how would that achieve the intended balance of objectives?

8. The following response stands upon the following assumptions:
    - As with any other industry, AI developers must pay for the supplies they use to build their products and deliver services;
    - The TDM exception is currently well-articulated, applying where computational analysis is conducted for the sole purpose of research for non-commercial use;
    - The concerns the consultation raises about a lack of legal clarity for AI developers is valid; and
    - The House of Lords' statement about copyright being for rewarding creatives must be enacted.

9. With this, we repeat that option 1 is the preferred path forward. However, it must be improved further. Firstly, it must be made clear that the spirit of section 29A of the *Copyright, Designs and Patents Act 1988* is upheld, as per the House of Lords' statement.

10. Secondly, creatives' work must not be available for machine learning training by default, as the opt-out option suggests. Protecting creative industries means, among other things, not burdening them. The need to negotiate licences should not fall to them but to the AI developers, who should seek ways to buy the rights to relevant works.

11. Thirdly, it should be clarified that option 1 can enable growth and innovations in licensing services, as the process of selling rights to an AI developer may be much easier through collective licensing and licensing bodies. With this, option 1 would improve with relevant provisions, reducing negative impacts on both creative and AI industries.

---

[6] Kairoi Ltd (2023) RE: Large language models inquiry, *Github*, online [accessed 01 January 2025]

[7] House of Lords Communications and Digital Committee (2024) Large language models and generative AI, *House of Lords*, online [accessed 01 January 2025]

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

12. First and foremost, justify why that body of work is necessary for the models they are working on. If they cannot justify the use of the particular copy of a work, they shouldn't pursue it further.

13. Second, see what information is made publicly available about the work's reservation: was it reserved by default, or does its creator or curator have a clear stance against their work being used for AI? In line with the previous response and assuming that works are reserved from their use for ML training by default, it is likely that many works are protected without any publicly articulated rationale from their creators. In such cases, it is appropriate for developers to contact the pertinent IP rights holders with a request for a change in their reservation status.

14. Consider a fictional example of a developer of an image-generating tool specialised in generating images of animals. The developer is aware that a certain zoo has extensive galleries of animal pictures. Following the UK's legislation in line with Kairoi's advice, the zoo's galleries remain online and under their IP by default. The developer approaches the zoo's leadership team (possibly the legal team) and makes a strong case for the usefulness of their own image-generating tool. The zoo then can decide whether or not they want to make their gallery available to the developer, or even remove the reservation entirely.

15. There is a strong chance that IP rights holders take clear, public stances about their decision to protect their creative outputs from ML use cases. Such stances may be found in press releases, blogposts or other communications. Where an IP rights holder takes a clear stance stance against their work being used for AI, the developer should not pursue its use any further.

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

16. The right holder should be compensated accordingly by the AI developer, much like with current intellectual property offences.[8]

Question: 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.

17. Yes. The basic workflow might look like datasets from AI models being audited and containing metadata about the licensing agreement for each data point. The machine readability of licensing metadata would make it inexpensive for the AI developer to adequately collect relevant data. Conversely, the cost for the creative individual or organisation would be to ensure their website(s) adequately integrates the .txt files that rescind or grant rights for AI developers. This cost may be minimal.

18. Finally, the greatest cost would be for the creation of the standard and its implementation. Here, it is worth noting that there are already .txt standards that have, notwithstanding,

---

[8] Intellectual Property Office (2017) Guidance: Intellectual property offences, online [accessed 01 January 2025]

been ignored by AI companies.[9] We suggest focusing on improving .txt standards and establishing fines for AI developers that do not respect relevant .txt files.

19. The cost here can be either for the AI developer to pay for AI audits, or for the IPO to have a specialised group that audits AI models for adherence to relevant .txt files. This may be in-line with the mention of collaborating across jurisdictions and industries.[10]

Technical Standards

Question: How can compliance with standards be encouraged?

20. This is the question we consider most important. At Kairoi, we help organisations develop more responsible AI cultures, and deem compliance a question of culture. The aforementioned case of AI companies bypassing .txt files is a sign of poor organisational cultures, as such files are a longstanding standard.

21. At a basic level, we can ensure compliance through a combination of incentives and penalties. Intellectual property offences are already fined, and AI-specific offences should be more clearly identified as such to continue safeguarding the rights of creatives. Fines deter offences and signal that certain acts are punishable.

22. Turning to the incentives, it is paramount that compliance with standards be deemed a social question rather than a technical one. Consider the Open Modelling Foundation's (OMF) learnings from convening diverse stakeholders of computational models to establish standards; including on AI. In short, the OMF found evidence that standards must be established collaboratively; involving different interested parties, acknowledging potential conflicts, and seeking ways to continually learn together and across disciplines.[11] In the case of this consultation, this might take the form of convening AI developers, policy makers, artists, musicians and so on to co-create standards that make sense to as many stakeholders as possible. This can help towards the consultation's objective regarding trust between the sectors.

23. The IPO has previously failed to create a voluntary code of practice with both AI and creative sectors stakeholders.[12] The magnitude of the challenge is precisely why this collaboration matters: an end-product that is agreed upon by both parties will undoubdtedly be extremely valuable. What we suggest here is a curated (and more transparent[13]) series of touchpoints between respresentatives of both stakeholders with a clear mandate from the IPO: we need clear standards to implement the letter and spirit of the law; to make it easy (i) for AI developers to know what data they can and cannot use to

---

[9] Paul, K. (2024) Exclusive: Multiple AI companies bypassing web standard to scrape publisher sites, licensing firm says, *Reuters*, online [accessed 01 January 2025]

[10] In §C.2 of the consultation

[11] Kherroubi Garcia, I. et al. (2025) Ten simple rules for good model-sharing practices, *PLOS Computational Biology*, 21(1), DOI: 10.1371/journal.pcbi.1012702

[12] DSIT (2024) A pro-innovation approach to AI regulation: government response, para. 29, online [accessed 01 Juanuary 2025]

[13] Trapova, A. (2024) UK's short-lived dream for a code of practice on genAI and copyright law, *Kluwer Copyright Blog*, online [accessed 19 January 2025]

train ML models; (ii) for creators to know how to communicate the IP of their creations to AI developers; and (iii) for the judicial system to hold creatives and AI developers to account.

24. The objectives of such a collaboration must be clear from the outset, and the objectives (as suggested in the previous paragraph) should be framed by the benefits each involved party will gain from the process.

25. The compromises that result can then inform standards, the implementation of which would require further collaboration. On the one hand, a campaign that appeals to creative industries may raise awareness about their rights and obligations. On the other hand, both industry and government may work on building dashboards and online portals where datasets are made available for their auditing and licensing management.

26. It will be key that adherence to the standard be technically and financially feasible. This may require greater investments at the outset to build the multidisciplinary community of practice and, eventually, the agreed-upon infrastructure.

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

27. Yes. As in our previous responses to consultations, we strongly believe the UK government has a role to play as convenor of relevant stakeholders and experts. Through this consultation, three governmental bodies are demonstrating their ability to convene and guide important discussions related to AI. What's more, in the particular area of legislation, it is important that the government ensure that the UK's values and interests are upheld throughout the process of standards co-creation.

28. As stated in the previous response, the role of the government should be to facilitate a cross-industry collaboration that is geared towards clarifying the rights and obligations of AI developers and creatives.

Transparency

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

29. Yes. This is for at least two reasons. Firstly, not disclosing training material sources enables AI developers to make false claims about the capabilities of their products. In one instance, Google's CEO claimed that an AI chatbot had somehow taught itself Bengali. This was proven to be false by Margaret Mitchell from Hugging Face; the training dataset underpinning the chatbot indeed had content in Bengali.[14]

30. Secondly, having to disclose the data sources for the training of AI tools can help AI developers foster more responsible tech cultures. Consider OpenAI's CTO who, in conversation with *The Wall Street Journal*'s Joanna Stern, claimed not to know whether

---

[14] Dixit, P. (2023) Researchers accused Google and "60 Minutes" of spreading AI "disinformation", *BuzzFeed*, online [accessed 02 January 2025]

their video-generating product was trained on videos from YouTube.[15] This was not a complex question, and YouTube is a well-known platform for watching videos. The question could have been foreseen by OpenAI staff, but the CTO appeared to be completely unprepared for it. Were it common practice to publicly disclose training data sources, we believe OpenAI would have a much more thoughtful approach to how they communicate about their products with the public. Indeed, better communications is one of the four pillars of responsible AI that we advocate for at Kairoi.[16]

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

31. It is crucial that the level of granularity not infringe on licensing agreements nor privacy law. For example, if an artist sells a license that uniquely grants one AI developer access to their artwork, then the AI developer should only disclose that some artwork (perhaps the number of art pieces) offered by a named or unnamed artist was present at the point of training their model. Similar practices might be instated in the case of data provided through a commercial contract. Conversely, if an AI model is trained on personal data, that data must not be disclosed. Rather, metadata should be disclosed: what is the nature of the data, how was it collected, when was it collected, how is it managed, who are the data processor and controller, and so on.

32. Given the aforementioned cases where Google and OpenAI executives appeared not to know about data sources, there is a need for a greater deal of granularity than there is at present.

Encouraging Research and Innovation

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

33. Yes, although modifications need at least clarify that the law applies to AI developers.

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

34. These are two very distinct factors. Whilst the size of an AI firm should not matter (startups must comply with the law and have obligations to preserve intellectual property rights as much as larger AI firms), an AI model's purpose may be relevant but warrants further investigation.

35. The distinction between commercial and non-commercial research purposes may be unclear in the context of AI. The gap between AI research and AI products is extremely blurry: non-commercial AI research may result in a commercial product quite straightforwardly. What's more, industry expertise may be required for non-commerical AI research in the first place.

---

[15] Stern, J. (2024) OpenAI made me crazy videos — then the CTO answered (most of) my questions, *The Wall Street Journal*, online [accessed 02 January 2025]

[16] Kherroubi Garcia, I. (2024) The four pillars of responsible AI, *Kairoi*, online [accessed 02 January 2025]

36. At Kairoi, we advocate for open science practices —evidenced by our ongoing work with the OMF and GO FAIR US.[17] With this, we view the need for transparency to transcend disciplinary and sectorial boundaries: universities, government, industry and the third sector at large must be transparent (within the aforementioned licensing and privacy constraints) about the data that underpins the AI models they develop. What's more, there is a need to be transparent about a model's purpose, as we propose with the OMF in *Ten Simple Rules for Good Model-sharing Practices*. However, the distinction in section 29A is difficult to uphold if we follow open science practices and work transparently, given the ease with which non-commercial research could result in commercial products.

37. To avoid burdening universities with having to negotiate for the licence of some relevant data for research-only purposes, it may be valuable to have a research-specific licence that grants free access to some data. Right holders could then sell a second licence to whatever firm decides to productise the university's research. This option may create a possible loophole whereby companies develop commercial products in partnership with universities. Hence, more investigation is needed.

Once again, we are grateful for the opportunity to respond to this consultation.

Ismael Kherroubi Garcia, FRSA
Founder and CEO, Kairoi
On behalf of Kairoi Ltd

---

[17] Kherroubi Garcia, I. (2024) 2024 in review, *Kairoi,* online [accessed 02 January 2025]