# Comments to the U.S. Copyright Office Inquiry on Copyright and Artificial Intelligence

**IT for Change**

**October 2023**

# IT for Change's Comments to the U.S. Copyright Office Inquiry on Copyright and Artificial Intelligence

## General Inputs

At the outset, IT for Change would like to appreciate the efforts of the United States Copyright Office (Copyright Office) towards understanding the intersection of artificial intelligence (AI) and copyrights. Against the backdrop of the revolutionizing developments in the field of generative AI (Gen AI), significant issues and challenges around the consumption and utilization of knowledge – both public and copyrighted – have come to the forefront.[1] The existing knowledge and understanding of fair use and copyright law are now being tested given the large amount of information – including copyrighted material – that is being fed to AI systems for training.[2]

Copyright law adopts a predominantly anthropocentric approach – copyright protection on written work, for instance, lapses 70 years after the author of the work dies.[3] Today, therefore, AI-generated works cannot obtain copyright protection as the current definition of personhood does not cover AI. Even assuming that in the future this definition of personhood is broadened, AI-produced works should not be granted copyright without examining the implications for the fair use of the intellectual and knowledge commons.

The *en masse* appropriation of text, visual, and audio material available in the public domain for the training of Gen AI models has raised concerns about creating knowledge enclosures. Open datasets are often used to build proprietary and closed systems. Some scholars have proposed that the evaluation of fair use in this scenario may require the introduction of a new ethical principle that one can term 'fair learning'. In this use, copyrighted input material can be used to train Gen AI models only when it can be ensured that the output of such AI systems does not pose "significant substitutive competition" to the authors/content creators whose works have been used. In this view, "If the purpose of the AI's use (of copyrighted input material) is not to obtain or incorporate the copyrightable elements of a work but to access, learn, and use the unprotectable parts of the work, that use should be presumptively fair."[4]

---

[1] United Nations Commission on Science and Technology for Development. (2023). *Issues Paper on Data for Development.* https://unctad.org/system/files/information-document/CSTD2023-2024_Issues01_data_en.pdf

[2] Ibid.

[3] Lucchi, N. (2023). *ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems.* Cambridge University Press. https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/chatgpt-a-case-study-on-copyright-challenges-for-generative-artificial-intelligence-systems/CEDCE34DED599CC4EB201289BB161965

[4] Lemley, M.A., & Casey, B. (2020). *Fair Learning.* Stanford Law School. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3528447

The violation of the moral rights of the author(s)[5] and cultural appropriation of indigenous art, language, and linguistic resources in the development of AI models[6] are other equally important concerns that should determine 'fair learning' exemptions.

As things stand, the legal and regulatory scheme is still miles away from addressing the concerns of the Gen AI ecosystem – one that the issue of copyrightability alone cannot expound.

## Summary of Recommendations

1. AI-produced works must be classified as uncopyrightable.

2. The permissibility of the use of copyrighted material for training AI models must be evaluated against a new ethical standard of 'fair learning' – ensuring the output of such AI systems does not pose 'significant substitutive competition' to the authors/content creators whose works have been used. In the case of collective knowledge, particularly traditional and indigenous knowledge, the risk of cultural appropriation and enclosure must also be a significant factor in the assessment of 'fair learning'.

3. Regulation must provide for the following minimum obligations:

   a. Mandate labeling of AI-generated work as distinct from human creations, disclosure of training data, specification of sources used for the same, and a clear, legible outline of the methodology of training.

   b. Build standard contract requirements that provide for standardized definitions (for example, 'prompt', in relation to the original idea entered into a Gen AI system to derive a response) and multidisciplinary collaboration. Standardization could be a stepping stone towards addressing bargaining power inequities by offering organizations a series of viable options for contractual rights, duties, and risks.[7]

   c. Establish collective licensing models where a Collective Management Organization (CMO) can negotiate terms of use and remuneration. However, the remuneration model adopted must ensure sufficient compensation to the concerned artists.[8]

---

[5] The Authors Guild. (2023). *Authors Guild Policy Proposals Regarding the Development and Use of Generative AI.* https://authorsguild.org/app/uploads/2023/07/Authors-Guild-AI-Policy-Proposals-7.19.23.pdf; See also, Federal Trade Commission. (2023). *Creative Economy and Generative AI.* https://www.ftc.gov/system/files/ftc_gov/pdf/creative-economy-and-generative-ai-transcript-october-4-2023.pdf; The Authors Guild. (n.d.). *Artificial Intelligence.* https://authorsguild.org/advocacy/artificial-intelligence/
[6] Hendrix, J. (2023). *An Indigenous Perspective on Generative AI.* Tech Policy Press. https://www.techpolicy.press/an-indigenous-perspective-on-generative-ai/
[7] When AI generates work, standard contractual terms can help generate value and clarity, see, https://oecd.ai/en/wonk/contractual-terms
[8] Shutterstock recently adopted a revenue-sharing model where "Contributors whose content was involved in training generative models will receive a share of the earnings from datasets and downloads of all AI-generated content produced on [their] platform […] Contributors will receive a share of the entire contract value paid by platform partners. The share individual contributors receive will be proportionate to the volume of their content and metadata that is included in the purchased datasets." Keeping in mind the extent of training data collected, it is likely that this compensation amounts to "fractions of a penny", as pointed out by Steven Zapata, from the

4.  The promotion and maintenance of open datasets should not create a 'free for all' mandate for data sharing. Protections outside of copyrights could be granted to ensure that open datasets do not allow for-profit entities to enclose public knowledge for private gain, thereby depleting the intellectual and knowledge commons.[9]

## Specific Inputs

**Q.1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?**

**Response:** The harms associated with Gen AI are synonymous with AI at large – the most significant of which is the rise in carbon emissions, the exploitation of labor, the reinforcement and reproduction of racial and gendered stereotypes and biases, and the inequitable distribution of social and economic benefits of AI-generated insights, with the value derived from such intelligence flowing primarily to very few large corporations in just two national jurisdictions – the US and China.[10] In keeping with the mandate of the U.S. Copyright Office, this response is limited to the following – the impact of Gen AI on indigenous artists; the context of authors having to compete with Gen AI outputs; and knowledge and value capture arising from the use of open-source datasets to build proprietary systems.

Training and data mining- (TDM) related activities serve as a critical step towards developing Gen AI models; the wider and more diverse the dataset, the more "accurate" the response. A crucial element of the training process includes the use of complete and exact replicas of original work, and this includes not just work protected by copyrights but also public domain knowledge that lack intellectual property protection, including, traditional knowledge commons. Gen AI systems are stochastic or *agere sine intelligence* in nature, i.e., "they act without understanding exactly what they are generating".[11] In effect, the Gen AI system simply creates a layer of noise in order to construct art or text. This, however, is a reconstruction of the art or text that is part of its training data.[12] A grant of copyright on Gen AI outputs, would, therefore, render indigenous artists without the means to exercise their intellectual

Concept Arts Association. See: Vincent, J. (2022). *Shutterstock will start selling AI-generated stock imagery with help from OpenAI.* The Verge. https://www.theverge.com/2022/10/25/23422359/shutterstock-ai-generated-art-openai-dall-e-partnership-contributors-fund-reimbursement

[9] Nowacki, I. (2023). *The Challenges of Digital Intelligence in the AI Revolution*. TS2. https://ts2.space/en/the-challenges-of-digital-intelligence-in-the-ai-revolution/

[10] PWC estimates that 70% of all economic value generated by AI will likely accrue to just two countries: the USA and China. *See*: Sudan, R., Hammer, C., & Eferin, Y. (2023). *Toward Bridging the Data Divide*. World Bank Blogs. https://blogs.worldbank.org/opendata/toward-bridging-data-divide

[11] Lucchi, N. (2023). *ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems.* Cambridge University Press. https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/chatgpt-a-case-study-on-copyright-challenges-for-generative-artificial-intelligence-systems/CEDCE34DED599CC4EB201289BB161965

[12] Hendrix, J. (2023). *An Indigenous Perspective on Generative AI.* Tech Policy Press. https://www.techpolicy.press/an-indigenous-perspective-on-generative-ai/

property rights, making them vulnerable to knowledge theft and misappropriation. This phenomenon is already prevalent in non-digital arenas such as in the case of Nike incorporating a traditional *Samonan* tattoo on a pair of leggings.[13,14] Additionally, this raises ethical considerations, specifically around new forms of colonization. For instance, the *Lakotas*, a polynesian community, are concerned about their "material objects that are going to be scanned through photogrammetry, or created, that are basically rip-offs of Indigenous cultures to be exploited".[15] It is likely that Gen AI will undercut efforts to preserve and maintain traditional knowledge practices, as it continues to be used as a knowledge production and management technology.

**Q.4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States?**

**Response:** Regulation around Gen AI is in its nascent stages. The first step towards regulating AI in the few jurisdictions where we see movement in the regulatory landscape is to impose transparency obligations around training data. Accountability measures in the development of Gen AI models are also recommended in the European Union Artificial Intelligence Act (EU AI Act).[16]

Transparency obligations must extend to the publication of training data, and a summary of training data used,[17] training methodology, and disclosures clearly distinguishing human and AI-generated work. The EU AI Act also emphasizes the need to set up responsive grievance redressal systems[18] that would help address harms caused by the use of Gen AI. Taking cognizance of the risk of Gen AI, the EU also provides for content moderation in order to ensure the training dataset does not contain any illegal material. This protection can also include ethical considerations to ensure diversity and accuracy of datasets so as to minimize bias.[19] Such a regulation must also consider a provision to ensure audit and oversight by a regulatory authority to independently verify the accuracy of datasets. Without effective oversight and monitoring mechanisms, it is likely that audit and risk management governance processes will fall short of the desired outcomes.

---

[13] Ibid.

[14] Vezina, B. (2019). *Curbing Cultural Appropriation in the Fashion Industry with Intellectual Property.* WIPO Magazine. https://www.wipo.int/wipo_magazine/en/2019/04/article_0002.html

[15] Hendrix, J. (2023). *An Indigenous Perspective on Generative AI.* Tech Policy Press. https://www.techpolicy.press/an-indigenous-perspective-on-generative-ai/

[16] Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on AI (AI Act) and amending certain Union legislative Acts. See, https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_E N.pdf

[17] Ibid.

[18] China Law Translate. (2023). *Article 15, Interim Measures for the Management of Generative Artificial Intelligence Services.* https://www.chinalawtranslate.com/en/generative-ai-interim/

[19] China Law Translate. (2023). *Article 4(2), Interim Measures for the Management of Generative Artificial Intelligence Services.* https://www.chinalawtranslate.com/en/generative-ai-interim/

The exemptions granted under the Single Digital Market in the European Union, 2019, allow TDM in relation to public data for commercial uses. To exercise their rights, copyright holders have to expressly exclude their works from such public datasets. Opt-out consent mechanisms, such as the one laid down by the European Union, are insufficient on two accounts. First, they shift the burden of awareness of the potential use of copyrighted material in training sets on copyright holders. This is untenable because those within the creative industry have to sift through large volumes of data in order to opt out or withdraw consent from their work being used by the relevant training model.[20] Secondly, an opt-out measure merely guarantees that the relevant work will not be used in 'future' training datasets. This provides no protection against existing AI models that have been trained on the relevant work.

In contrast to the EU model, an opt-in mechanism is built on a 'permission first' ethical principle; one that has witnessed effective use in other sectors. Reflecting this principle, China's Interim Administrative Measures for Generative Artificial Intelligence Services, 2023, provide for compliance with intellectual property rights at the input (with regards to training data) and output (with regards to works produced by Gen AI)[21] stages. Further, they stipulate that express consent must be provided for the inclusion of personal data in the training dataset.[22]

**Q.9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?**

**Response:** The scope of licensing exceptions to the use of copyrighted works in training AI models must widen beyond an over simplified distinction between the commercial and non-commercial use of data. A familiar term in data markets is that of 'sugging', i.e., building datasets under the guise of conducting research. The loopholes created by a classification as broad as this have been exploited to a large extent through fake reviews where companies may hire people to fill out reviews or by getting consumers to fill out surveys to receive a prize. The first step, therefore, must be to develop a stricter definition of commercial use of publicly available data.

Secondly, fostering the creation and maintenance of open datasets must include robust regulation. The identification and preservation of traditional knowledge systems, for instance, requires stringent limitations regarding use. Such a regulation could ensure that the social and economic value generated from such work is distributed fairly to the creators of work –including those who may not hold the bundle of intellectual property rights, but have rights of ownership over their work through other regulatory mechanisms, such as contractual agreements with larger corporations that allow work on a

---

[20] Federal Trade Commission. (2023). *Creative Economy and Generative AI.* https://www.ftc.gov/system/files/ftc_gov/pdf/creative-economy-and-generative-ai-transcript-october-4-2023.pdf
[21] China Law Translate. (2023). *Article 4(3), Interim Measures for the Management of Generative Artificial Intelligence Services.* https://www.chinalawtranslate.com/en/generative-ai-interim/
[22] China Law Translate. (2023). *Article 7(3), Interim Measures for the Management of Generative Artificial Intelligence Services.* https://www.chinalawtranslate.com/en/generative-ai-interim/

made-for-hire basis. A similar principle was crystallized in the Indian Arts and Crafts Act (IACA) of 1990, which mandates accurate labeling of indigenous artwork so as to reduce consumer deception and preserve traditional knowledge.

**Q.10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?**

**Q.10.2. Is a voluntary collective licensing scheme a feasible or desirable approach?**

**Response:** A collective licensing scheme would be a step towards recognizing the cultural commons of creative work and heritage. This must strike a delicate balance between the protection of the intellectual and cultural commons and fostering innovation in technology. In the absence of specific regulation, a collective license could limit the reuse of work in violation of the cultural commons. A CMO could negotiate licenses for a specific class of work (such as images) for a specified category of workers, regardless of whether they are members of the CMO. The license terms will allow creators (rather than corporate copyright owners) to retain the copyright. This is particularly useful in sectors where artists sign over their rights to the agency they work with.[23]

The U.S. Copyright Office may, as provided by the Authors Guild,[24] be authorized to:

- Authorize organizations to function as CMOs who represent classes of work and rights holders, and

- Provide for a mechanism, such as a Copyright Royalty Board, to mediate licensing rates, should free market negotiations fail.

The licensing agreements may provide for the following:

**Consent Mechanisms**

- Notice should be sent by the CMO so as to allow the artists to exercise the right to opt out of the CMO;

- Notice should be sent for opt-in to training datasets at large (although this may be negotiated by the CMO), and

- The Copyrights Office may provide the terms of notice.

---

[23] The Authors Guild. (2023). *Authors Guild Policy Proposals Regarding the Development and Use of Generative AI.* https://authorsguild.org/app/uploads/2023/07/Authors-Guild-AI-Policy-Proposals-7.19.23.pdf; See also, Federal Trade Commission. (2023). *Creative Economy and Generative AI.* https://www.ftc.gov/system/files/ftc_gov/pdf/creative-economy-and-generative-ai-transcript-october-4-2023.pdf
[24] Ibid.

**Scope of Use**

- Similar to purpose limitation in data protection regulations, scope of use clauses can define and restrict the use of data. For instance, 3-D body scans of models cannot be used by companies to tick the diversity check box.

Such a licensing mechanism will not be inclusive or effective in isolation. It is important that it is backed by regulation across different areas, especially labor and antitrust laws.[25] In order to ease licensing negotiations, the Protect Working Musicians Act, 2023, was introduced in Congress earlier this year.[26] It aims to support musicians through licensing negotiations, without antitrust laws prohibiting collective decision-making.

---

[25] Ibid.

[26] Congress member Deborah Ross introduced proposals to empower indie artists in digital deals. See, https://ross.house.gov/2023/9/federal-lawmaker-introduces-revamped-protect-working-musicians-act-proposing-indie-collective-negotiation-rights-for-streaming-and-generative-ai