



## UNITED STATES COPYRIGHT OFFICE

# Artificial Intelligence and Copyright

**PUBLISHED ON WEDNESDAY, AUGUST 30, 2023**

**59942 FEDERAL REGISTER, VOL. 88, NO. 167**

**ACTION:** Notice of inquiry and request for comments.

**SUMMARY:** The United States Copyright Office is undertaking a study of the copyright law and policy issues raised by artificial intelligence (“AI”) systems. To inform the Office’s study and help assess whether legislative or regulatory steps in this area are warranted, the Office seeks comment on these issues, including those involved in the use of copyrighted works to train AI models, the appropriate levels of transparency and disclosure with respect to the use of copyrighted works, and the legal status of AI-generated outputs.

**DATES:** Written comments are due no later than 11:59 p.m. Eastern Time on Wednesday, October 18, 2023. Written reply comments are due no later than 11:59 p.m. Eastern Time on Wednesday, November 15, 2023.

**ADDRESSES:** For reasons of governmental efficiency, the Copyright Office is using the [regulations.gov](https://www.regulations.gov) system for the submission and posting of public comments in this proceeding. All comments should be submitted electronically through [regulations.gov](https://www.regulations.gov). Specific instructions for submitting comments are available on the Copyright Office website at <https://copyright.gov/policy/artificial-intelligence>. If electronic submission is not feasible, please contact the Office using the contact information below for special instructions.

**FOR FURTHER INFORMATION CONTACT:** Rhea Efthimiadis, Assistant to the General Counsel, by email at [meft@copyright.gov](mailto:meft@copyright.gov) or telephone at 202-707-8350.

## SUPPLEMENTARY INFORMATION

### I. Introduction

Over the last year, artificial intelligence (“AI”) systems and the rapid growth of their capabilities have attracted significant media and public attention. One type of AI, “generative AI” technology, is capable of producing outputs such as text, images, video, or audio (including emulating a human voice) that would be considered copyrightable if created by a human author.<sup>1</sup> The adoption and use of generative AI systems by millions of Americans<sup>2</sup>—and the resulting volume of AI-generated material—have sparked widespread

---

1. Generative AI technologies produce outputs based on “learning” statistical patterns in existing data, which may include copyrighted works. Kim Martineau, *What is generative AI?*, IBM Research Blog (Apr. 20, 2023), <https://research.ibm.com/blog/what-is-generative-ai> (“At a high level, generative models encode a simplified representation of their training data and draw from it to create a new work that’s similar, but not identical, to the original data.”). The Office has defined “generative AI” and other key terms in a glossary at the end of this Notice.

2. See, e.g., *Microsoft FY23 Second Quarter Earnings Conference Call Transcript*, Microsoft (Jan. 24, 2023), <https://www.microsoft.com/en-us/Investor/events/FY-2023/earnings-fy-2023-q2.aspx> (Microsoft CEO Satya Nadella stating that “[m]ore than one million people have used Copilot to date”); Krystal Hu, *ChatGPT sets record for fastest-growing user base - analyst note*, Reuters (Feb. 2, 2023), <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.

public debate about what these systems may mean for the future of creative industries and raise significant questions for the copyright system.<sup>3</sup>

Some of these questions relate to the scope and level of human authorship, if any, in copyright claims for material produced in whole or in part by generative AI. Over the past several years, the Office has begun to receive applications to register works containing AI-generated material, some of which name AI systems as an author or co-author.<sup>4</sup> At the same time, copyright owners have brought infringement claims against AI companies based on the training process for, and outputs derived from, generative AI systems.<sup>5</sup> As concerns and uncertainties mount, Congress and the Copyright Office have been contacted by many stakeholders with diverse views. The Office has publicly announced a broad initiative earlier this year to explore these issues. This Notice is part of that initiative and builds on the Office’s research, expertise, and prior work, as well as information that stakeholders have provided to the Office.

## II. The Copyright Office’s Past Work on Machine Learning and AI

The Copyright Office has long been engaged in questions involving machine learning and copyright. In 1965, the Office’s annual report noted that developments in computer technology had begun to raise “difficult questions of authorship”—namely the question of the authorship of works “‘written’ by computers.”<sup>6</sup> As the then-Register of Copyrights observed:

---

3. See, e.g., James Vincent, *The scary truth about AI copyright is nobody knows what will happen next*, The Verge (Nov. 15, 2022), <https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data> (discussing the “key [legal] questions from which the topic’s many uncertainties unfold”); see Kevin Roose & Cade Metz, *How to Become an Expert on A.I.*, N.Y. Times (Apr. 4, 2023), <https://www.nytimes.com/article/ai-artificial-intelligence-chatbot.html>; Kim Martineau, *What is generative AI?*, IBM Research Blog (Apr. 20, 2023), <https://research.ibm.com/blog/what-is-generative-ai>; Harvard Online, *The Benefits and Limitations of Generative AI: Harvard Experts Answer Your Questions*, Harvard Online Blog (Apr. 19, 2023), <https://www.harvardonline.harvard.edu/blog/benefits-limitations-generative-ai>; Arhan Islam, *A History of Generative AI: From GAN to GPT-4*, Marktechpost (Mar. 21, 2023), <https://www.marktechpost.com/2023/03/21/a-history-of-generative-ai-from-gan-to-gpt-4/>. Generative AI is also a point of contention in the labor disputes between the Alliance of Motion Picture and Television Producers and both the Writers Guild of America and SAG-AFTRA (the guild representing actors and other media professionals). See Andrew Webster, *Actors say Hollywood studios want their AI replicas – for free, forever*, The Verge (July 13, 2023), <https://www.theverge.com/2023/7/13/23794224/sag-aftra-actors-strike-ai-image-rights>.

4. See U.S. Copyright Office Review Board, *Decision Affirming Refusal of Registration of A Recent Entrance to Paradise* at 2 (Feb. 14, 2022), <https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf> (noting visual work was submitted listing the author as the “Creativity Machine”).

5. See, e.g., Am. Compl. ¶¶ 8, 61, *Getty Images (US), Inc. v. Stability AI, Inc.*, No. 1:23-cv-135, ECF No. 13 (D. Del. Mar. 29, 2023) (alleging infringement based on use of copyrighted images to train a generative AI model and on the possibility of that model generating images “highly similar to and derivative of” copyrighted images).

6. U.S. Copyright Office, *Sixty-Eighth Annual Report of the Register of Copyrights for the Fiscal Year Ending June 30, 1965*, at 5 (1966), <https://www.copyright.gov/reports/annual/archive/ar-1965.pdf>.

The crucial question appears to be whether the “work” is basically one of human authorship, with the computer merely being an assisting instrument, or whether the traditional elements of authorship in the work (literary, artistic, or musical expression or elements of selection, arrangement, etc.) were actually conceived and executed not by man but by a machine.<sup>7</sup>

Because the answer depends on the circumstances of a work’s creation, the head of the Office’s Examining Division (and future Register) Barbara Ringer warned that the Office could not “take the categorical position that registration will be denied merely because a computer may have been used in some manner in creating the work.”<sup>8</sup> As she noted, “a typewriter is a machine that is used in the creation of a manuscript[,] but this does not result in the manuscript being uncopyrightable.”<sup>9</sup> This view was echoed a decade later by the National Commission on New Technological Uses of Copyrighted Works (“CONTU”),<sup>10</sup> which agreed with the Office<sup>11</sup> but declined to discuss the issue in depth because “[t]he development of this capacity for ‘artificial intelligence’ has not yet come to pass, and, indeed, it has been suggested to this Commission that such a development is too speculative to consider at this time.”<sup>12</sup> In the intervening years, as AI moved out of the realm of speculation, the Office continued to participate in discussions on AI issues, from a 1991 conference hosted by the World Intellectual Property Organization (“WIPO”)<sup>13</sup> to more recent events the Office co-hosted with WIPO<sup>14</sup> and with the U.S. Patent and Trademark Office.<sup>15</sup>

---

7. *Id.*

8. U.S. Copyright Office, *Annual Report of the Examining Division, Copyright Office, for the Fiscal Year 1965*, at 4 (1965), <https://copyright.gov/reports/annual/archive/ar-examining1965.pdf>.

9. *Id.*

10. CONTU was created “to assist the President and Congress in developing a national policy for both protecting the rights of copyright owners and ensuring public access to copyrighted works when they are used in computer and machine duplication systems.” CONTU, *Final Report of the National Commission on New Technological Uses of Copyrighted Works* at 3 (July 31, 1978) (“CONTU Final Report”) One of its statutory mandates was to study “the creation of new works by the application or intervention of [] automatic systems or machine reproduction.” *National Commission on New Technological Uses of Copyrighted Works*, Pub. L. 93-573, sec. 201(b)(2), 88 Stat. 1873 (1974).

11. CONTU Final Report at 44–46 (recommending the same “approach [that] is followed by the Copyright Office today in conducting examinations for determining registrability for copyright of works created with the assistance of computers”).

12. *Id.* at 44.

13. See U.S. Copyright Office, *94th Annual Report of the Register of Copyrights for the Fiscal Year Ending September 30, 1991*, at 2 (1991), <https://copyright.gov/reports/annual/archive/ar-1991.pdf>.

14. See *Copyright in the Age of Artificial Intelligence*, U.S. Copyright Office (Feb. 5, 2020), <https://www.copyright.gov/events/artificial-intelligence/>.

15. See *Copyright law and machine learning for AI: Where are we and where are we going?*, U.S. Patent and Trademark Office (Oct. 26, 2021), <https://www.uspto.gov/about-us/events/copyright-law-and-machine-learning-ai-where-are-we-and-where-are-we-going>. The Office also supported the U.S. Patent and Trademark Office when it solicited public comments on the impact of AI on intellectual property policy, including copyright. See U.S. Patent and Trademark Office, *Public Views on Artificial Intelligence and Intellectual Property Policy* (Oct. 2020), [https://www.uspto.gov/sites/default/files/documents/USPTO\\_AI-Report\\_2020-10-07.pdf](https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf).

Last year, in two separate copyright registration matters, the Office publicly addressed the question of copyright in AI-generated material. In the first instance, the Office refused to register a claim for two-dimensional artwork described as “autonomously created by a computer algorithm running on a machine.”<sup>16</sup> The Office’s Review Board<sup>17</sup> explained that the work could not be registered because it was made “without any creative input or intervention from a human author,” and that “statutory text, judicial precedent, and longstanding Copyright Office practice” all require human authorship as a condition of copyrightability.<sup>18</sup> The Office’s registration denial, as well as the supporting legal analysis, was recently affirmed in federal district court.<sup>19</sup>

A second registration application, submitted in 2022, involved a work containing both human authorship and generative AI material. The work was a graphic novel with text written by the human applicant and illustrations created through the use of Midjourney, a generative AI system. After soliciting information from the applicant about the process of the work’s creation, the Office determined that copyright protected both the human-authored text and human selection and arrangement of the text and images, but not the AI-generated images themselves.<sup>20</sup> The Office explained that where a human author lacks sufficient creative control over the AI-generated components of a work, the human is not the “author” of those components for copyright purposes.<sup>21</sup> The Office continues to receive applications to register works incorporating AI-generated material, involving different levels of human contributions.<sup>22</sup>

---

**16.** U.S. Copyright Office Review Board, *Decision Affirming Refusal of Registration of A Recent Entrance to Paradise* at 2 (Feb. 14, 2022), <https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf>.

**17.** The Review Board is a three-member body that hears administrative appeals of copyright registration decisions. Review Board decisions constitute final agency actions and are subject to judicial review. *See* 37 CFR 202.5(f), (g).

**18.** U.S. Copyright Office Review Board, *Decision Affirming Refusal of Registration of A Recent Entrance to Paradise* at 3 (Feb. 14, 2022), <https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf>.

**19.** Mem. Op., *Thaler v. Perlmutter*, No. 22-cv-1564, ECF No. 24 (D.D.C. Aug. 18, 2023).

**20.** U.S. Copyright Office, *Cancellation Decision re: Zarya of the Dawn (VAu001480196)* at 1 (Feb. 21, 2023), <https://www.copyright.gov/docs/zarya-of-the-dawn.pdf> (letter from the Office to applicant canceling the original certificate and issuing a new one covering only the expressive material created by the applicant).

**21.** *Id.* at 9.

**22.** In addition to registration, the Office has considered AI in the regulatory context of the section 1201 rulemaking. Section 1201 of the Copyright Act sets up a triennial proceeding to address possible exceptions to a statutory ban on circumventing technological protection measures that control access to copyrighted works. *See* 17 U.S.C. 1201(a)(1)(C) (charging Register of Copyrights with making recommendation as to whether particular users of copyrighted works are adversely affected in ability to engage in noninfringing uses). In the most recent proceeding, the Register was asked to consider text and data mining activities as part of this analysis, and she concluded that existing copyright case law did not support the conclusion that all such activity is fair use. The Register did, however, recommend granting a narrow exemption after concluding that the specific use as described was likely to be fair because it was limited to a “researcher or group of researchers seeking to investigate a particular set of questions that require examination of a large number of works;” access to the works in full would be limited to researchers solely for purposes of verifying research results; and the researchers would not use the works “for their expressive purposes.” U.S. Copyright Office, *Section 1201 Rulemaking: Eighth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention, Recommendation of the Register of Copyrights* 107–13 (Oct. 2021).

### III. The Office’s AI Initiative

In response to growing Congressional<sup>23</sup> and public interest,<sup>24</sup> the Office launched a comprehensive AI Initiative in early 2023. The Initiative identified a number of steps that the Office would take to further explore the copyright policy questions surrounding AI, including hosting public listening sessions and publishing a notice of inquiry.<sup>25</sup> At the same time, the Office created a website, [www.copyright.gov/ai](http://www.copyright.gov/ai), to provide information about the Initiative, including planned events and opportunities for public engagement.

#### a. March 2023 Registration Guidance

At the outset of the Initiative, the Office issued a statement of policy providing registration guidance on works containing AI-generated material (“AI Registration Guidance”).<sup>26</sup> The AI Registration Guidance reiterated the principle that copyright protection in the United States requires human authorship. Under well-established case law, the Guidance explained, “the term ‘author,’ used in both the Constitution and the Copyright Act, excludes non-humans.”<sup>27</sup> In the context of generative AI, this means that “[i]f a work’s traditional elements of authorship were produced by a machine, the work lacks human authorship and the Office will not register it.”<sup>28</sup> The Guidance instructed applicants seeking to register works containing more than *de minimis* AI-generated material to disclose that the work contains such material and provide a brief explanation of the human author’s contributions.<sup>29</sup>

#### b. Public Listening Sessions

In April and May 2023, the Office held four public listening sessions to gather input on the copyright issues raised by generative AI. Each session focused on a different

---

23. See Letter from Sen. Chris Coons, Chair, and Sen. Thom Tillis, Ranking Member, Subcomm. on Intell. Prop. of the S. Comm. on the Judiciary, to Kathi Vidal, Under Secretary of Commerce for Intell. Prop. and Director, U.S. Patent and Trademark Office, and Shira Perlmutter, Register of Copyrights, U.S. Copyright Office (Oct. 27, 2022) and Letter from Kathi Vidal, Under Secretary of Commerce for Intell. Prop. and Director, U.S. Patent and Trademark Office, and Shira Perlmutter, Register of Copyrights, to Sen. Chris Coons, Chair, and Sen. Thom Tillis, Ranking Member, Subcomm. on Intell. Prop. of the S. Comm. on the Judiciary (Dec. 12, 2022), <https://www.copyright.gov/laws/hearings/Letter-to-USPTO-USCO-on-National-Commission-on-AI-1.pdf> (Senate letter requesting the Office to provide guidance on what the law around generative AI should be in the future and the Office’s response explaining that it intended, among other things, to issue a notice of inquiry on questions involving copyright and AI).

24. See, e.g., *Virtual AI Townhall hosted by Karla Ortiz featuring the US Copyright Office*, Concept Art Ass’n (Nov. 2, 2022), <https://www.conceptartassociation.com/calendar/virtual-ai-townhall-featuring-us-copyright-office> (event that featured two senior attorneys from the Office).

25. *Copyright Office Launches New Artificial Intelligence Initiative*, U.S. Copyright Office (Mar. 16, 2023), <https://www.copyright.gov/newsnet/2023/1004.html>.

26. *Copyright Registration Guidance: Works Containing Materials Generated by Artificial Intelligence*, 88 FR 16190 (Mar. 16, 2023). A copy of the guidance is available at [https://copyright.gov/ai/ai\\_policy\\_guidance.pdf](https://copyright.gov/ai/ai_policy_guidance.pdf).

27. *Id.* at 16191.

28. *Id.* at 16192.

29. *Id.* at 16193.

category of creative work: literary works, including print journalism and software; works of visual art; audiovisual works, including video games; and musical works and sound recordings. Over the four listening sessions, nearly 90 participants representing individual artists, academic experts, legal practitioners, technology companies, and industry associations shared their views with the Office. Transcripts, videos recordings, and agendas for each session are available on the Office’s website.<sup>30</sup>

*c. Educational Webinars*

In June and July 2023, the Office held two public webinars on generative AI, each of which drew an audience of nearly 2,000. The first webinar focused on registration of works containing AI-generated material. It included an overview of the Office’s general rules on how to register works containing material created or owned by someone other than the applicant, followed by examples illustrating how those rules apply to works that incorporate AI-generated material.<sup>31</sup> The second webinar convened experts on different regions of the world to discuss international developments in generative AI and copyright law. These experts discussed how other countries are addressing copyright issues, including authorship, training, and exceptions and limitations. They provided an overview of legislative developments and highlighted possible areas of convergence and divergence.<sup>32</sup>

*d. Engagement with Stakeholders*

In addition to the public events described above, the Office has spoken with a broad spectrum of stakeholders, participating in dozens of meetings with academics, trade groups, individual creators, technology companies, and creative industries.<sup>33</sup> These meetings have provided valuable information on the technical aspects of generative AI models and systems, how creators are using generative AI, and the continuing questions copyright applicants have about registering works that include AI-generated material.

---

**30.** *Spring 2023 AI Listening Sessions*, U.S. Copyright Office, <https://www.copyright.gov/ai/listening-sessions.html>.

**31.** The transcript and recording of the registration webinar are available at <https://www.copyright.gov/events/ai-application-process/>. In the coming months, the Office intends to provide further guidance to copyright applicants seeking to register works containing AI-generated material.

**32.** The transcript and recording of the international webinar are available at <https://www.copyright.gov/events/international-ai-copyright-webinar/>.

**33.** Additionally, the Office has offered guidance to The Mechanical Licensing Collective (“The MLC”), explaining that AI-generated music is not eligible for the statutory mechanical blanket license in section 115 of the Copyright Act and that The MLC should not disburse royalties for such musical works. See Letter from Suzanne V. Wilson, General Counsel and Associate Register of Copyrights, U.S. Copyright Office, to Kris Ahrend, Chief Exec. Officer, The MLC, at 2–3 (Apr. 20, 2023), <https://www.copyright.gov/ai/USCO-Guidance-Letter-to-The-MLC-Letter-on-AI-Created-Works.pdf>.

#### IV. The Current Inquiry

Drawing on our prior AI Initiative work, including discussions with stakeholders, the Office has identified a wide range of copyright policy issues arising from the development and use of AI. These relate to: (1) the use of copyrighted works to train AI models; (2) the copyrightability of material generated using AI systems; (3) potential liability for infringing works generated using AI systems; and (4) the treatment of generative AI outputs that imitate the identity or style of human artists. The Office seeks public comments on these and related issues.

As to the first issue, the Office is aware that there is disagreement about whether or when the use of copyrighted works to develop datasets for training AI models (in both generative and non-generative systems) is infringing.<sup>34</sup> This Notice seeks information about the collection and curation of AI datasets, how those datasets are used to train AI models, the sources of materials ingested into training, and whether permission by and/or compensation for copyright owners is or should be required when their works are included. To the extent that commenters believe such permission and/or compensation is necessary, the Office seeks their views on what kind of remuneration system(s) might be feasible and effective. The Office also seeks information regarding the retention of records necessary to identify underlying training materials and the availability of this information to copyright owners and others.

On the second issue, the Office seeks comment on the proper scope of copyright protection for material created using generative AI. Although we believe the law is clear that copyright protection in the United States is limited to works of human authorship,<sup>35</sup> questions remain about where and how to draw the line between human creation and AI-generated content. For example, are there circumstances where a human's use of a generative AI system could involve sufficient control over the technology, such as through the selection of training materials and multiple iterations of instructions ("prompts"), to result in output that is human-authored? Resolution of this question will affect future registration decisions. While the Office is separately working to update its registration guidance on works that include AI-generated material,<sup>36</sup> this Notice explores the broader policy questions related to copyrightability.

---

<sup>34</sup> In some cases, a non-generative AI model may be trained on copyrighted material. In other cases, the same AI model may be capable of being deployed in both a generative AI system and a non-generative one. The Office's consideration of training is framed broadly in order to encompass these and other situations.

<sup>35</sup> See Mem. Op., *Thaler v. Perlmutter*, No. 22-cv-1564, ECF No. 24 (D.D.C. Aug. 18, 2023) (affirming the Office's registration denial of AI-generated work).

<sup>36</sup> For example, the Office has received questions about how to apply its guidance that applicants disclose more than *de minimis* amounts of AI-generated material in their works. See AI Registration Guidance, 88 FR at 16193 (explaining that "AI-generated content that is more than *de minimis* should be explicitly excluded from the application").

On the third question, the Office is interested in how copyright liability principles could apply to material created by generative AI systems.<sup>37</sup> For example, if an output is found to be substantially similar to a copyrighted work that was part of the training dataset, and the use does not qualify as fair, how should liability be apportioned between the user whose instructions prompted the output and developers of the system and dataset?

Lastly, in both our listening sessions and other outreach, the Office heard from artists and performers concerned about generative AI systems' ability to mimic their voices, likenesses, or styles. Although these personal attributes are not generally protected by copyright law, their copying may implicate varying state rights of publicity and unfair competition law, as well as have relevance to various international treaty obligations.<sup>38</sup>

## V. Overview of Notice

The purpose of this Notice is to collect factual information and views relevant to the copyright law and policy issues raised by recent advances in generative AI. The Office undertakes this study pursuant to its statutory mandate in title 17 to “[c]onduct studies” and “[a]dvise Congress on national and international issues relating to copyright, other matters arising under this title, and related matters.”<sup>39</sup> It intends to use this information to advise Congress by providing analyses of the current state of the law, identifying unresolved issues, and evaluating potential areas for congressional action. The Office will also use this record to inform its regulatory work and to offer information and resources to the public, courts, and other government entities considering these issues.

The questions are grouped into several categories. This Notice begins with several general high-level questions and then inquires about AI training, including questions of transparency and accountability; generative AI outputs, including questions of copyrightability, infringement, and labeling or identification of such outputs; and other issues related to copyright. Because of the importance of using shared language in discussing these issues, the questions are followed by a glossary of key terms for the purposes of this Notice. The Office welcomes input from commenters on the definitions.

---

<sup>37</sup> Some of these questions are currently before the courts in lawsuits that have already been filed over generative AI systems. See, e.g., *J.L. v. Alphabet Inc.*, 3:23-cv-03340 (N.D. Cal.); *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-3417 (N.D. Cal.); *Silverman v. OpenAI, Inc.*, 4:23-cv-3416 (N.D. Cal.); *Tremblay v. OpenAI, Inc.*, 3:23-cv-3223 (N.D. Cal.); *Getty Images (US), Inc. v. Stability AI, Inc.*, 1:23-cv-0135 (D. Del.); *Andersen v. Stability AI Ltd.*, 3:23-cv-0201 (N.D. Cal.); *Doe v. GitHub, Inc.*, 4:22-cv-6823 (N.D. Cal.).

<sup>38</sup> See U.S. Copyright Office, *Authors, Attribution, and Integrity: Examining Moral Rights in the United States* 112–116 (Apr. 2019), <https://www.copyright.gov/policy/moralrights/full-report.pdf> (discussing how such interests are generally protected under state right of publicity laws).

<sup>39</sup> 17 U.S.C. 701(b)(1), (b)(4).



## VI. Instructions and Questions

The Office does not expect that every party choosing to respond to this Notice will address every question raised below. The questions are designed to gather views from a broad range of parties. The Office does request that, when responding to a question, commenters clearly identify each question for which they submit a response, address questions separately, and provide the factual, legal, or policy basis for their responses. Commenters should make clear whether they are submitting in a personal capacity or on behalf of an organization or entity they are authorized to represent. Commenters are particularly encouraged to explain any technical understandings that inform their legal and policy viewpoints, as well as whether their answers are applicable only to certain industries, technologies, or types of copyrighted works. Although some questions seek technical information about generative AI systems, commenters do not need to be affiliated with a technical entity to answer these questions.

### General Questions:

*The Office has several general questions about generative AI in addition to the specific topics listed below. Commenters are encouraged to raise any positions or views that are not elicited by the more detailed questions further below.*

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?
2. Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?
3. Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.
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?<sup>40</sup> How important a factor is international consistency in this area across borders?

---

<sup>40</sup> For example, several jurisdictions have adopted copyright exceptions for text and data mining that could permit use of copyrighted material to train AI systems. Separately, the European Parliament passed its version of the Artificial Intelligence Act on June 14, 2023, which includes a requirement that providers of generative AI systems publish “a sufficiently detailed summary of the use of training data protected under copyright law.” See Artificial Intelligence Act, amend. 399, art. 28b(4)(c), EUR. PARL. DOC. P9\_TA (2023)0236 (2023), [https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html).

5. Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

### Training

*If your comment applies only to a specific subset of AI technologies, please make that clear.*

6. What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?
  - 6.1. How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?
  - 6.2. To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?
  - 6.3. To what extent is non-copyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?
  - 6.4. Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.
7. To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:
  - 7.1. How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.
  - 7.2. How are inferences gained from the training process stored or represented within an AI model?
  - 7.3. Is it possible for an AI model to “unlearn” inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to “unlearn” inferences from training?
  - 7.4. Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?

8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.
  - 8.1. In light of the Supreme Court's recent decisions in *Google v. Oracle America*<sup>41</sup> and *Andy Warhol Foundation v. Goldsmith*,<sup>42</sup> how should the "purpose and character" of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning,<sup>43</sup> raise different considerations under the first fair use factor?
  - 8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?
  - 8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes.<sup>44</sup> How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature?<sup>45</sup> Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?
  - 8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?
  - 8.5. Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured?<sup>46</sup> Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?
9. Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

---

**41.** 141 S. Ct. 1183 (2021).

**42.** 143 S. Ct. 1258 (2023).

**43.** See *Pre-training, Fine-tuning, and Foundation Models*, GenLaw: Glossary (June 1, 2023), <https://genlaw.github.io/glossary.html> (explaining that pre-training is a relatively slow and expensive process that "results in a general-purpose or foundation model" whereas fine-tuning "adapts a pretrained model checkpoint to perform a desired task using additional data").

**44.** For example, the generative AI model, Stable Diffusion, was reportedly developed in part by researchers at the Ludwig Maximilian University of Munich but is used by the for-profit company Stability AI. See Kenrick Cai, *Startup Behind AI Image Generator Stable Diffusion Is In Talks To Raise At A Valuation Up To \$1 Billion*, *Forbes* (Sept. 7, 2022), <https://www.forbes.com/sites/kenrickcai/2022/09/07/stability-ai-funding-round-1-billion-valuation-stable-diffusion-text-to-image/?sh=31e11f5a24d6>.

**45.** 17 U.S.C. 107(1).

**46.** *Id.* at 107(4).

- 9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?<sup>47</sup>
- 9.2. If an “opt out” approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?<sup>48</sup>
- 9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?
- 9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?
- 9.5. In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire—should they have a right to object to an AI model being trained on their work? If so, how would such a system work?
10. If copyright owners’ consent is required to train generative AI models, how can or should licenses be obtained?
  - 10.1. Is direct voluntary licensing feasible in some or all creative sectors?
  - 10.2. Is a voluntary collective licensing scheme a feasible or desirable approach?<sup>49</sup> Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress

---

**47.** For example, the European Union’s Directive on Copyright in the Digital Single Market provides for two copyright exceptions or limitations for text and data mining (which may be used in the training of generative AI systems): one for purposes of scientific research and one for any other purpose. The latter is available only to the extent that rightsholders have not expressly reserved their rights to the use of their works in text and data mining. See Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 2019 O.J. (L 130), <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

**48.** For example, some AI companies have reportedly started to allow copyright owners to tag their works as not available for AI training. See Emilia David, *Now you can block OpenAI’s web crawler*, The Verge (Aug. 7, 2023), <https://www.theverge.com/2023/8/7/23823046/openai-data-scrape-block-ai>; Melissa Heikkilä, *Artists can now opt out of the next version of Stable Diffusion*, MIT Tech. Review (Dec. 16, 2022), <https://www.technologyreview.com/2022/12/16/1065247/artists-can-now-opt-out-of-the-next-version-of-stable-diffusion/>.

**49.** Collective licensing is one alternative to a direct licensing regime, in which copyright owners negotiate and enter into private agreements on an individual basis. Under a collective licensing arrangement, rights are aggregated and administered by a management organization. The management organization negotiates the terms of use and distributes payment to participating copyright owners. See WIPO, *WIPO Good Practice Toolkit for CMOs* at 6 (2021), <https://www.wipo.int/publications/en/details.jsp?id=4561>.

consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

- 10.3. Should Congress consider establishing a compulsory licensing regime?<sup>50</sup> If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?
- 10.4. Is an extended collective licensing scheme<sup>51</sup> a feasible or desirable approach?
- 10.5. Should licensing regimes vary based on the type of work at issue?
11. What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?
12. Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system? Please explain.
13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?
14. Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.

#### Transparency & Recordkeeping

15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

---

**50.** A compulsory or “statutory” license allows for certain uses of a copyrighted work “without the consent of the copyright owner provided that the person adhered to the provisions of the license, most notably paying a statutorily established royalty to the copyright owner.” *Music Licensing Reform: Hearing Before the Subcomm. on Intell. Prop. of the S. Comm. on the Judiciary*, 109th Cong. (2005) (statement of Marybeth Peters, Register of Copyrights), <http://copyright.gov/docs/regstat071205.html>.

**51.** “An Extended Collective Licensing scheme is one where a relevant licensing body, subject to certain safeguards, is authori[z]ed to license specified copyright works on behalf of all rights holders in its sector (including non-members), and not just members who have given specific permission for it to act.” Extended Collective Licensing (ECL) scheme definition, LexisNexis Glossary (2023), <https://www.lexisnexis.co.uk/legal/glossary/extended-collective-licensing-ecl-scheme>; see also Letter from Karyn A. Temple, Acting Register of Copyrights, U.S. Copyright Office, to Rep. Robert Goodlatte, Chair, and Rep. John Conyers, Ranking Member, H. Comm. on the Judiciary (Sept. 29, 2017), <https://www.copyright.gov/policy/massdigitization/house-letter.pdf>; Letter from Karyn A. Temple, Acting Register of Copyrights, U.S. Copyright Office, to Sen. Charles Grassley, Chair, and Sen. Dianne Feinstein, Ranking Member, S. Comm. on the Judiciary (Sept. 29, 2017), <https://www.copyright.gov/policy/massdigitization/senate-letter.pdf>.

- 15.1. What level of specificity should be required?
- 15.2. To whom should disclosures be made?
- 15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?
- 15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?
16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?
17. Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?

### **Generative AI Outputs**

*If your comment applies only to a particular subset of generative AI technologies, please make that clear.*

### **Copyrightability**

18. Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?
19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?
20. Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?
  - 20.1. If you believe protection is desirable, should it be a form of copyright or a separate *sui generis* right? If the latter, in what respects should protection for AI-generated material differ from copyright?

21. Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection “promote the progress of science and useful arts”?<sup>52</sup> If so, how?

### Infringement

22. Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?
23. Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?
24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?
25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?
  - 25.1. Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs?<sup>53</sup>
26. If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the system?
27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.

### Labeling or Identification

28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?
  - 28.1. Who should be responsible for identifying a work as AI-generated?

---

<sup>52</sup>. U.S. Const. art. I, sec. 8, cl. 8.

<sup>53</sup>. Some AI models are released by their developers for download and use by members of the general public. Such so-called “open-source” models may restrict how those models can be used through the terms of a licensing agreement. *See, e.g., Llama 2 Community License Agreement, Meta AI (July 18, 2023), <https://ai.meta.com/llama/license/>* (requiring users of Llama 2 AI model to include an attribution notice and excluding use in services with greater than 700 million monthly active users).

- 28.2. Are there technical or practical barriers to labeling or identification requirements?
- 28.3. If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?
- 29. What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?

Additional Questions About Issues Related to Copyright

- 30. What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?
- 31. Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?
- 32. Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works “in the style of” a specific artist)? Who should be eligible for such protection? What form should it take?
- 33. With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws?<sup>54</sup> Does this issue require legislative attention in the context of generative AI?
- 34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

**VII. Glossary of Key Terms**

The Office has included definitions of key terms as they are used in this Notice to clarify the technical processes involved in generative AI systems. The following definitions are used for purposes of this Notice only; they do not necessarily reflect the government’s legal position with respect to any particular term.

*Artificial Intelligence (AI)*: A general classification of automated systems designed to perform tasks typically associated with human intelligence or cognitive functions.<sup>55</sup>

---

<sup>54</sup>. Under 17 U.S.C. 114(b), the reproduction and derivative work rights for sound recordings “do not extend to the making or duplication of another sound recording that consists entirely of an independent fixation of other sounds, even though such sounds imitate or simulate those in the copyrighted sound recording.”

<sup>55</sup>. See *John S. McCain National Defense Authorization Act for Fiscal Year 2019*, Pub. L. 115–232, sec. 238(g) (2), 132 Stat. 1636, 1697–98 (2018) (defining “artificial intelligence” to include systems “developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action”).



Generally, AI technologies may use different techniques to accomplish such tasks. This Notice uses the term “AI” in a more limited sense to refer to technologies that employ machine learning, a technique further defined below.

*AI Model:* A combination of computer code and numerical values (or “weights,” which is defined below) that is designed to accomplish a specified task. For example, an AI model may be designed to predict the next word or word fragment in a body of text. Examples of AI models include GPT-4, Stable Diffusion, and LLaMA.

*AI System:* A software product or service that substantially incorporates one or more AI models and is designed for use by an end-user.<sup>56</sup> An AI system may be created by a developer of an AI model, or it may incorporate one or more AI models developed by third parties.

*Generative AI:* An application of AI used to generate outputs in the form of expressive material such as text, images, audio, or video. Generative AI systems may take commands or instructions from a human user, which are sometimes called “prompts.” Examples of generative AI systems include Midjourney, OpenAI’s ChatGPT, and Google’s Bard.

*Machine Learning:* A technique for building AI systems that is characterized by the ability to automatically learn and improve on the basis of data or experience, without relying on explicitly programmed rules.<sup>57</sup> Machine learning involves ingesting and analyzing materials such as quantitative data or text and obtain inferences about qualities of those materials and using those inferences to accomplish a specific task. These inferences are represented within an AI model’s weights.

*Training Datasets:* A collection of training material (as defined below) that is compiled and curated for use in machine learning. Examples of training datasets include BookCorpus, ImageNet, and LAION.

*Training Material:* Individual units of material that are used for purposes of training an AI model. They may include a combination of text, images, audio, or other categories of expressive material, as well as annotations describing the material. An example of training material would be an individual image and an associated text “label” that describes the image.

*Weights:* A collection of numerical values that define the behavior of an AI model. Weights are stored within an AI model and reflect inferences from the training process.

---

<sup>56</sup>. See James M. Inhofe *National Defense Authorization Act for Fiscal Year 2023*, Pub. L. 117–263, sec. 7223(4) (A), 136 Stat. 2395, 3669 (2022) (defining “artificial intelligence system” as “any data system, software, application, tool, or utility that operates in whole or in part using dynamic or static machine learning algorithms or other forms of artificial intelligence”).

<sup>57</sup>. See *National Artificial Intelligence Initiative Act of 2020*, 15 U.S.C. 9401(11).

**DATED: AUGUST 24, 2023**

**SUZANNE V. WILSON, GENERAL COUNSEL AND ASSOCIATE REGISTER OF COPYRIGHTS.**

**MARIA STRONG, ASSOCIATE REGISTER OF COPYRIGHTS AND DIRECTOR OF POLICY AND  
INTERNATIONAL AFFAIRS.**

[FR DOC. 2023-18624 FILED 8-29-23; 8:45 AM]