

**Before the  
COPYRIGHT OFFICE,  
LIBRARY OF CONGRESS  
Washington, DC**

---

Notice of Inquiry: Artificial  
Intelligence and Copyright  
(FR Doc. 2023-18624;  
Copyright Office Docket  
Number 2023-6)

)  
)  
)  
) **COMMENTS OF THE**  
) **NATIONAL WRITERS UNION (NWU)**  
)  
)

---

**October 30, 2023**

**National Writers Union**  
25 Broadway, Room 1504  
New York, NY 10004  
212-254-0279  
[nwu@nwu.org](mailto:nwu@nwu.org)  
<<https://nwu.org>>

The National Writers Union (“NWU”) submits these comments in response to the Notice of Inquiry (“NOI”) and request for comments by the U.S. Copyright Office, “Artificial Intelligence and Copyright,” FR Doc. 2023-18624, Copyright Office Docket Number 2023-6, 88 *Federal Register* 59942-59949 (August 30, 2023). By notice promulgated on September 21, 2023 (86 *Federal Register* 65205, FR Doc. 2023-20480), the deadline for comments in response to this notice of inquiry was extended through October 30, 2023.

### **About the NWU and our interest in this proceeding**

The NWU is an independent national labor union that advocates for freelance and contract writers and media workers, including print and digital self-publishers. The NWU includes local chapters as well as at-large members nationwide and abroad. The NWU works to advance the economic conditions of writers and media workers in all genres, media, and formats.

NWU membership includes, among others, journalists, fiction and nonfiction book authors, poets, novelists, playwrights, editors, academic writers, business and technical writers, website and email newsletter content providers, bloggers, social media producers, podcasters, videographers, illustrators, photographers, graphic artists, translators, and other digital media workers. The NWU is a member of international federations including the International Federation of Journalists (“IFJ”), which represents 600,000 media professionals from 187 labor unions and associations in more than 140 countries.<sup>1</sup>

---

<sup>1</sup> “About IFJ: The Global Voice of Journalists”, <<https://www.ifj.org/who/about-ifj.html>>. The NWU also urges the Copyright Office and Congress to keep in mind the impact of U.S. copyright law and regulations on creative workers worldwide.

## **Introduction and overview**

As creators of “born digital” works distributed primarily online, freelance media workers are among those most likely to be affected by generative AI technologies. Writers, photographers, animators, illustrators, graphic designers, podcasters, editors, multimedia journalists—all of us, our work and our livelihoods, are on the line.

The work we create has been and continues to be used to “train” these systems without our knowledge or consent. Our careers, already precarious and devalued, are increasingly under threat, as corporations turn to generative AI as a “cost-effective” silver bullet. It’s no coincidence that the threat of generative AI is emerging at a time when media outlets are struggling to turn a profit, tech and media workers are being hit with wave after wave of layoffs, and union membership across the board is at a historic low.

Companies developing generative AI engines largely obscure details of how their technologies work—even if they are themselves able to understand those details in the first place. As such, when a creative worker’s copyrighted work shows up in a given dataset, it is nearly impossible for us as individuals to seek recompense. And those who wish to remove our work from these datasets have to rely on tools created by reverse engineering (e.g., Have I Been Trained?) to try to determine exactly which texts or images might have been used to train a given AI system.

But discovery isn’t enough: Even if a creative worker were able to identify where their work had been used, and to what extent, to produce a given generative AI output, that same creative worker would have to lean on officially registered copyrights to pursue any

compensation for that work having been used—a registration process whose fees often preclude this being a reasonable course of action for the average freelancer. What’s more, those of us who don’t want our work to be included in training corpuses at all face just as pitched a battle. At present, the scope of the web scraping that’s already been done to train the generative AI systems active today—which has been, in short, exhaustive—means that it is more than simply burdensome for any creative worker to effectively “opt out” of having our work be used as training data in the first place: It’s next to impossible.

We respond below to those questions in the Notice of Inquiry which we believe are most relevant to our concerns as creative workers. Our responses to these questions are informed by our Platform and Principles for Policy on Generative AI, as developed and ratified by our members. Our Platform and Principles for Policy on Generative AI include additional suggested action items for Congress and the Copyright Office related to copyright, antitrust, and labor law, as well as issues to be addressed through contracts and collective bargaining. We have attached our Platform and Principles for Policy on Generative AI as part of these comments, and commend them to Congress, the Copyright Office, and other agencies to guide their actions on this issue.

## **Responses to questions in the Notice of Inquiry**

***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?*

Unregulated use of generative AI without proper attribution will provide few benefits, all of which are outweighed by deeply impactful risks that will compromise human innovation.

**Benefits:**

- Creators may automate some of our workflow processes.

**Risks:**

- Unverified sources and black-box models lead to increased misinformation among text-based AIs.
- Private citizens and public figures alike are vulnerable to "deep fake" AI-generated voices, images, and videos, which may compromise their credibility and further spread misinformation, distrust, and conspiracy.
- Human progress is stymied when AI-generated content replaces the role of creative workers.
- Increased publication of undisclosed AI-generated content creates a feedback loop where AI models train on false information generated by AI, eroding the truth and diminishing generative AI capabilities.
- Risk of displacing human creative workers, either in part (creative tasks currently assigned to human workers getting “reassigned” to AI systems) or in whole (entire gigs being shifted from human workers to AI). Some NWU members are already seeing jobs they would have been commissioned for (illustrations, blurb-writing,

photography gigs) being produced by generative AI, while others are seeing a precipitous drop in proffered per-word rates, explicitly framed as being a pay rate based on the expectation that the writer will be using AI as a supplementary tool.

- Risk of misidentification of human creative works as AI-generated, if AI output isn't required to be clearly labeled as such (or vice versa, but both need to be recognized as potential problems). Some NWU members are already being asked to “prove” that their work is human, and denied freelance contracts if they can't do so.
- Risk of violating economic rights of creators of works used for “training.”
- Risk of violating moral rights of creators of works used for “training.”
- Risks of facilitating the production of defamation, spam, fake news, etc.

We are already witnessing an unprecedented barrage of false information driven by generative AI. Additionally, creative workers have lost their livelihoods due to the unregulated use of generative AI. These issues will compound without prompt intervention.

***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?***

Generative AI can mimic the style of any copyright owner or content creator whose works were used in the training corpus. These original creators are specialists who have extensively trained in their respective styles, amounting to significant temporal and financial

investments. Generative AI does not compensate these specialists in any way, either directly or indirectly.

Additionally, as a union that includes a significant number of journalist members, and which is a member of the IFJ, we note here that the risks inherent to generative AI-produced output with regards to plagiarism—not to mention fact-checking and implicit biases—are of significant concern for our industry.

***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.***

See the NWU’s Platform and Principles for Policy on Generative AI, included as an attachment to these comments, which provide a framework to guide policy-making on this issue.

***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? How important a factor is international consistency in this area across borders?***

Global efforts to introduce regulations relating to copyright and generative AI models currently focus on defining copyright protections for AI *outputs*. The United States should lead by example in protecting copyright owners and creative workers whose content has been used as *input* for generative AI “training” purposes.

The international norms that should be applied to AI include those established by existing international treaties to which the US is a party, including the Berne Convention. There are independent reasons for adopting legislation to address these issues—including moral rights, formalities, and permissible exceptions and limitations—but the problems with generative AI heighten the importance and urgency for the US to fully implement the unimplemented provisions of the Berne Convention.

International consistency is important, since without it AI developers will exploit any jurisdiction of convenience that provides a safe haven for AI “training” without permission or payment. The World Wide Web can be scraped from servers anywhere in the world. So it is important for the US to uphold international norms set by the Berne Convention and other treaties, and to insist that all other countries do so, including insisting that the EU close the loopholes created by the application to generative AI of the exceptions to copyright for so-called “text and data mining” in the 2019 EU Directive on Copyright in the Digital Single Market.

***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.***

We believe that new legislation is necessary to address copyright and other issues with generative AI. This legislation must protect the intellectual property of creators, guaranteeing creators credit, control, and compensation for content used in generative AI “training.”

Our policy recommendations, and the basis for them, are detailed and explained in our Platform and Principles for Policy on Generative AI, as attached to these comments.



These policies are rooted in the principles of solidarity, humanity, control of and compensation for our work, transparency, accountability, and integrity.

As described in the attached platform, we call for legislation to provide for:

### **1. Control**

Among other points, we believe that creators should never be replaced by generative AI. If generative AI is used, it should be as a tool to assist human workers and augment our creative work, not as a replacement. It should be workers, not employers, who determine what is and is not assistive and augmentative. This technology should always supplement, never supplant creative work.

### **2. Compensation**

Among other points, we believe that creators should be compensated for all work used for AI at every stage, that any fair compensation strategy must ensure that creators are paid an appropriate rate for our work, and that use for AI development should not be deemed to fall under “fair use” or similar exceptions to copyright. (That is, that Congress should explicitly exclude use for “training” of generative AI from the statutory definition of “fair use.”)

### **3. Credit, Labeling and Transparency**

Among other points, we believe that any work that was derived in whole or in part by generative AI should include credit (“attribution”) to the authors whose original creations were used to train the AI used to generate that work, that work

generated in whole or in part by AI should be labeled to protect audiences from work that is misleading or incorrect, and that all AI-generated works should include **both** credit and labeling, not just one or the other.

#### **4. Fair Contracts**

Among other points, we believe that future contracts should explicitly ask creators for permission to use our work in AI systems, with terms specifying exactly which generative AI systems a work will be used for, and not simply be a blanket agreement for any and all uses. We also believe that AI development was not anticipated or paid for in past contracts (including work-for-hire contracts or contracts in which “all” rights were sold and licensed). As such, these contracts should not be read to allow for use of these works in AI training. Congress should clarify in the definition of works created for hire, that rights to use for AI development are retained by human creators, even with respect to works created for hire, unless those rights were *explicitly* assigned in writing.

In addition, we believe that freelancers and self-publishers should be afforded the right to bargain collectively with publishers, platforms/distributors, and AI companies without fear of violating antitrust law. An antitrust exemption for creators of intellectual property could be modeled on the longstanding antitrust exemption for agricultural cooperatives.

#### **5. Compliance with the Berne Convention and other copyright treaties**

Among other points, we believe that Congress must enact legislation explicitly protecting and providing effective redress for violations of authors' moral rights, independent of copyright ownership or any mechanisms for enforcement of economic rights such as civil litigation by rightsholders or criminal prosecutions for copyright infringement.

***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?***

AI companies have not provided full credit to creators or comprehensive indexes of the works used for training. This has frustrated impact assessments, redress (if we don't know our works have been used, we will have difficulty establishing standing to sue for copyright infringement or violation of licensing terms requiring attribution in derivative works), and assertion of creators' moral rights to attribution and objection to prejudicial use.

Generative AI Model training processes and the materials used for training purposes are proprietary, and this information is mostly unavailable to the public. In the absence of transparency by AI companies, we have been forced to rely on investigative reporting and

reverse engineering to try to find out which works by which authors have been copied. Most AI, especially for text, appears to have been trained primarily with unauthorized mirror copies of Web content. According to OpenAI, ChatGPT3 was trained on 45TB of text data, which is the equivalent of 95 million novels. There is no public registry of this training data nor transparency regarding the licensing of copyrighted works.

***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?***

OpenAI and similar generative AI model developers claim to use some public domain works in training. Still, all content creators and their estates deserve to be notified and protected by prospective legislation, regardless of whether or not they hold the official copyright.

AI companies have also claimed that they have used “open source” content for training, by which they presumably mean works licensed under Creative Commons or other “open source” licenses. But all Creative Commons licenses and almost all other open source licenses require attribution and a link to the license. This is intended to respect the rights of creators to credit, and to enable verification of compliance with licensing terms. Use of these allegedly “licensed” works for AI training has been and is being done in violation of these license terms.

***7. To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained.***

Many NWU members have found our works on reverse-engineered indexes of Web sites, books, and images used for training specific AI models

***8. Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use?***

We don't believe that it should be considered fair use. Fair use is a statutory creation, and if necessary, Congress should clarify this in the statutory definition of fair use.

***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)?***

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

***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?***

The fairest and most equitable solution to gathering data for AI training is through an opt-in basis so content creators of all technical abilities may retain the right to refuse to license their work for AI training purposes. An opt-out approach is not a feasible option for some

creative workers and copyright owners. Tools like technical flags and metadata can be prohibitive for those unfamiliar with digital technologies and people with impairments that impact their ability to utilize these tools.

Work by freelancers is typically published by third parties in formats and on Web sites for which freelance creators of text and other content elements do not control the metadata.

For previously licensed or published works, freelance creators typically have no contractual authority to demand that publishers add or modify metadata.

Moreover, we have already seen international examples of individual artists being countersued for legal costs when pursuing the removal of their work from training sets that they didn't "opt-in" to. To make an "opt-out"-only approach precedent in U.S. law would enable similar unjust (and financially untenable) outcomes for any creative workers in the United States who might be interested in removing their work from LLM datasets.

***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?***

An opt-in process can be established by generative AI model developers and/or third parties responsible for training dataset cultivation, creating a marketplace where creators and copyright owners can sell licenses to extant work. For Web content, an opt-in tag could easily be included on Web pages. Furthermore, this marketplace would facilitate more job opportunities for creative workers to produce niche content for AI training.

***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?***

Existing civil remedies are inadequate, because (1) litigation is effectively unavailable for unregistered works, including Web content that can't affordably be registered, because attorneys' fees can't be recovered for infringement of unregistered works, and (2) there are currently no remedies under US law for violations of moral rights of authors of written works.

***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?***

Moral rights are independent of copyright ownership, so the Berne Convention requires the US to create a remedy for violations of moral rights independent of copyright ownership, including for the human creators of works made for hire.

Since generative AI can mimic the style of creative workers, consent must be required from not only the copyright owner but also the original creator of works used to train AI models. Every human must retain the explicit right to withhold their creative work from AI training datasets, and this right must be legally enforceable.

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

Which works to license, for which uses, to which users, at what price, and on what terms are all issues that should be determined through collective bargaining between creators and would-be users. Organizing of creative workers and collective bargaining to determine the answers to these licensing questions cannot be conducted on a free and fair basis until creative workers, including freelancers and self-publishers, are free to organize and bargain collectively without fear of antitrust enforcement.

Once creative workers are free to organize and bargain collectively without fear of antitrust enforcement, a marketplace where creators and copyright owners can sell licenses to extant work should be developed to resolve this issue. A marketplace like this would facilitate more job opportunities for creative workers to produce niche content for AI training.

Additionally, as stated in our attached platform regarding the development of a collective licensing scheme to compensate creators, such a scheme might be supported by a reproduction rights organization (RRO). In that case, any RRO carrying out collective licensing on behalf of creative workers should be a member-governed creator organization, on the model of e.g. worker co-ops or producer co-ops.

***10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? 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 consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?***

Voluntary licensing could be one of the many ways in which generative AI corporations are held accountable for their use of creative works that do not belong to them.

But existing organizations are chilled by fear of possible antitrust enforcement, which impedes efforts to organize creative workers into collective licensing organizations. Congress should grant an antitrust exception for producers of intellectual property, perhaps modeled on the longstanding and successful antitrust exemption for agricultural producer cooperatives to allow creators to negotiate the terms of these collective licensing agreements and engage in other joint marketing activities.

***10.4. Is an extended collective licensing scheme a feasible or desirable approach?***

A prerequisite for an extended collective licensing (“ECL”) scheme is a sufficiently representative organization of creators of the category of works to be licensed. But most Web content creators are not members of any organization, and organizing of freelancers and self-publishers in all genres and media is deterred by fear of antitrust action. No ECL should be considered until an unambiguous antitrust exemption is created to permit organizing by freelance and self-published creators.

***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)?***

There are no legal, technical, or practical issues, current or foreseen, to obtaining appropriate licensing for training data. A suitable model is already in use across the digital publishing industry that generative AI developers and third parties involved may adopt. Consider how many millions of Web publishers voluntarily opted in to display advertising on their sites, in exchange for compensation, once attractive compensation was available for doing so.

Creators can currently sell their content on stock marketplaces (e.g., Shutterstock), where they can set attributes for acceptable use, pricing, and licensing terms. Companies and individuals can browse these marketplaces to shop for photos, videos, audio tracks, etc., and purchase the appropriate licenses. Generative AI developers and third parties involved in training datasets may follow this established approach and disclose their intent to use licensed content for training AI. Curators of AI training datasets should be responsible for securing appropriate licensing for content, and developers training AI models must verify that licensing is accurate and current.

***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.***

The technology for identifying the sources involved in a generative AI output is limited. However, efforts are being made to ascertain the composition of Generative AI training corpora, including the *Washington Post's* analysis of “The websites in Google’s C4 dataset” for Web content (<https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/>), *The*

*Atlantic's* analysis of the Books3 database

(<https://www.theatlantic.com/technology/archive/2023/08/books3-ai-meta-llama-pirated-books/675063/>) and “Have I Been Trained” (<https://haveibeentrained.com/>) for images.

***13. What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?***

The economic impact of implementing opt-in licensing requirements for generative AI system training data would be a net positive. It would stimulate the economy significantly by providing marketplaces with ample opportunities for creative workers to license existing works and produce new content for training purposes. Additionally, requirements for licensing, attribution, citations, and disclosures would improve the quality of AI-generated content.

***14. Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models.***

As stated in our attached platform, when it comes to journalism, credibility is of critical importance. It is of enormous concern to the NWU that AI can be used to generate voices or images that can be nearly impossible for people to distinguish as fabrications, or text that cites false or—most directly relevant to this question—erroneously plagiarized information. If left unchecked—e.g., if systems aren’t put in place to ensure that any given output will contain *zero* plagiarized material—the consequences will be dire.

***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?***

It is customary among comparable industries and academia to cite sources for factual content gleaned from extant publications. Generative AI model developers and any third parties involved in cultivating training datasets should adhere to this standard. They should look to citation guidelines established by the APA, MLA, or CMOS for specificity and format.

***15.1. What level of specificity should be required?***

The simplest and easiest way to provide attribution would be through a hyperlink from each AI output to a searchable online index by author and identifier (URL, ISCC, ISBN, etc.) to each work included in the corpus of training material. This would be similar to the routine use of hyperlinks from credit lines to details of Creative Commons or other open source licensing terms.

***15.2. To whom should disclosures be made?***

Disclosures and citations must be made public-facing and accessible to all. Furthermore, each prompt output must provide the end-user with a source breakdown to properly credit original creators and enable fact-checking to counter problematic or erroneous output.

***15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?***

Developers of AI systems that incorporate models from third parties must be responsible for verifying that all licensing information is accurate and updated.

***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?***

There is no significant cost associated with reasonable recordkeeping and attribution efforts. Generative AI developers and third parties must follow the same protocols that comparable industries and academia already adhere to when citing sources. The most significant impact anticipated is that recordkeeping and attribution will improve output quality; i.e., cohesive and correct outputs can be traced to the copyrighted source material for further reference, while misleading and ungrammatical sources can be avoided or purged.

***16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?***

Developers and third parties involved in AI training must take reasonable measures to notify and compensate copyright owners whose works have already been used for AI training purposes.

***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?***

Generative AI end-users can use prompt engineering to access paywalled content and copyrighted material not publicly available.

***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?***

As addressed in our attached platform, we believe that users and providers of AI systems and services are responsible for their use. That is, that it is not only the end users who have a duty to consider how their use of these systems impacts the human creators behind the training corpora, as well as their own final audiences. It is also the developers' responsibility to ensure that their systems fairly compensate and credit the human creators whose work the systems rely on, and that the outputs of these systems meet high standards of accuracy and ethical integrity. All this is true regardless of who is deemed legally liable for the output of generative AI.

***25.1. Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs? (53)***

Creators must receive compensation and attribution for their work no matter which model is used. And as noted above, almost all “open source” licenses require attribution and a link to the license, which AI developers and service producers aren’t including in AI output.

***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?***

The law should require generative AI models to provide attribution and citations with every output. As stated in our attached platform:

- Any work that was derived in whole or in part by generative AI should include credit (“attribution”) to the authors whose original creations were used to train the AI used to generate that work.
- Credit entails a link to an online index that details all ingested work in the corpus (e.g. by identifiers such as an ISBN, URL, ISCC, etc.) used by the generative AI system(s) involved in the generation of the work.
- Work generated in whole or in part by AI should be labeled to protect audiences from work that is misleading or incorrect. Labeling entails some kind of mark, as appropriate for the medium in question, that indicates to the readers, viewers, or listeners that something has been generated with the help of AI.

- All AI-generated works should include both credit and labeling, not just one or the other. Together they serve two main purposes. The first is to give audiences insight into how something was made. The second is to allow creators to see where their work has been used to generate content. Without both of these tools, we won't be able to assert our economic and moral rights.

***28.1. Who should be responsible for identifying a work as AI-generated?***

The onus is on the developers of AI models to implement output attribution links to a publicly accessible, searchable index of all of the works in the training corpus.

***28.2. Are there technical or practical barriers to labeling or identification requirements?***

Any anticipated barriers that AI developers may encounter are not unreasonable or without precedent. Compelled labels are legally required in many sectors, including movie and TV ratings, safety labels on medications, nutritional labels, etc, as well as in academia.

***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?***

While many products have been sold with the promise of successfully identifying AI-generated content, they are inaccurate and often unreliable. OpenAI itself discontinued their AI detector due to inaccuracy on 20 July 2023



(<https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>). Other AI detectors still in use are prone to false positives and have adversely impacted writers and students who did not use generative AI in their processes.

***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? (54) Does this issue require legislative attention in the context of generative AI?***

As our attached platform underscores, we believe that all creative workers are due robust respect to their moral rights. Expanded right of publicity laws, statutory protection for authors' moral rights, and mandatory AI crediting would be steps in the right direction for audio workers as much as for any other creative worker.

***34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.***

Legislation to address the rights of creative workers with respect to generative AI needs to address issues of antitrust and labor law as well as copyright. We encourage the Copyright Office and Congress to take a holistic approach to generative AI legislation that recognizes the interaction of these areas of law, rather than a piecemeal approach.

To help provide guidance and recommendations for comprehensive generative AI legislation to address creative workers' rights, including issues not raised in the Notice of Inquiry from the Copyright Office, we have included the NWU's Platform and Principles for Policy on Generative AI as an attachment to these comments.

Our Platform and Policies on Generative AI have been developed and ratified by our membership through extensive discussion and a democratic and participatory process. They reflect the interests of our diverse membership and a concern for the rights of all creative workers, including those who are not members of the NWU or any organization.

We thank the Copyright Office for the opportunity to provide these comments. We look forward to working with the Copyright Office and Congress on legislation to address the issues that have led to this policy study.

Respectfully submitted,

**National Writers Union**

Larry Goldbetter, President

Edward Hasbrouck, NWU Generative AI Working Group

Rose Eveleth, NWU Generative AI Working Group

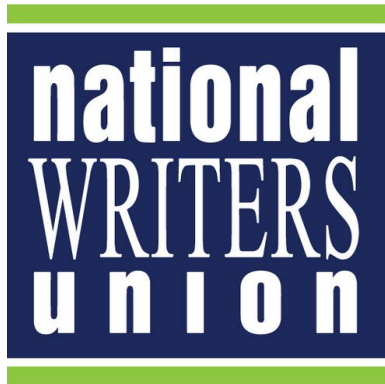
Alexis Gunderson, NWU Digital Media Division

Max Loel, NWU Generative AI Working Group

Phuc Pham, NWU Generative AI Working Group

& other members of the NWU Generative AI Working Group

Attachment: National Writers Union Platform and Principles for Policy on Generative AI



National Writers Union  
25 Broadway, Room 1504  
New York, NY 10004

212-254-0279  
nwu@nwu.org  
[www.nwu.org](http://www.nwu.org)

# National Writers Union

## Platform and Principles for Policy on Generative AI

### Introduction

In the past few years, the capabilities of technologies known as “**generative AI**” (see the [Glossary](#) below for defined terms in **blue**) have progressed by leaps and bounds. Across a wide range of media (writing, visual art, audio, video) these AI systems have become increasingly effective at generating content that is indistinguishable from that produced by human writers and creators.

As a union of professional creative workers, the membership of NWU has been watching these developments with interest and concern. This document, drafted over the summer and fall of 2023, is the culmination of conversations across the union and with collaborators, experts, and partner groups in the US and from around the world about the power and peril of generative AI. What follows is both our philosophy towards generative AI and our official union policy platform regarding what should be done to protect the lives, livelihoods, and labor of creators.

**Note:** *This platform is a living document and subject to change to reflect the position of our membership and as the landscape of AI shifts. Updates will be discussed by the NWU Generative AI Working Group and vetted by the NWU membership before going live. Contact [Rose](#), [Alexis](#), or [Edward](#) if you aren't currently part of the working group and would like to get involved. This version 1.0 was adopted and ratified in October 2023.*

# PRINCIPLES

The Generative AI platform of the National Writers Union (NWU) is informed by the following principles:

## Solidarity

***We believe that engaging with the broad spectrum of impacts of generative AI on workers and societies around the world is crucial to the development of policies for generative AI.***

We endeavor to find solutions that benefit all kinds of creators, and protect creative work of the past, present and future. The National Writers Union comprises a broad spectrum of creators — from Web content writers, to photographers, podcasters, book authors, journalists and more. As generative AI technologies expand to displace human creators of almost every type of copyrighted **work**, we must remember not to leave any creative worker behind. Our work on this issue must be inclusive, participatory and democratic.

We must also acknowledge the disparate impacts of AI technologies on marginalized communities and workers who might not fall within our membership. More so than many technologies, generative AI reproduces, and sometimes enhances, pre-existing social inequities and biases such as racism, sexism, homophobia, transphobia and more. For example, AI systems “trained” primarily on works from the global North will generate output – even for AI users in the global South – that reflects Northern biases toward the South. Similarly, in order to produce higher quality results, the companies behind these AI systems exploit large numbers of so-called “ghost workers,” many also in the global South, to clean and categorize data. Many of these workers are not compensated fairly or treated humanely.

We also recognize that generative AI has huge implications for a broad range of workers, and believe there is an urgent need for both governments and society to recognize the need for a **just transition** for all workers. Where we cannot protect jobs from displacement by AI, we must ensure that we’re providing pathways to safe, just, and accessible economic opportunities.

National and regional governments must also resist pressure to be drawn into a race to the bottom to create a corporate-friendly, creator-unfriendly environment for AI development, as companies search the world for jurisdictions of convenience. Operating “in the cloud,” AI companies can locate their servers anywhere in the world. Currently, the US and the European Union (EU) offer the most favorable legal climate for AI development, so there is little incentive for AI companies to look further afield. But if the US and EU close the loopholes provided by broad interpretations of exceptions to copyright in their laws, AI companies will look for other countries where they can obtain more favorable laws.

Finally, we recognize that these technologies have outsized impacts on the climate through energy and water use. In addition to harming necessary efforts to mitigate the climate crisis, this will again disproportionately impact marginalized communities, often in ways that are largely invisible to most AI users.

## Humanity

***We believe that the rights of creators are human rights.***

Human creators are not the same as publishers, technologies, or corporations, none of which should be deemed to have “human” rights. People, corporations, and algorithms should not be lumped together or treated the same way. Generative AI works because it “**ingests**” voluminous amounts of human-made creativity – the work of millions of human lives – which should be protected from exploitation and erosion. On this issue, even more than in other copyright debates, our humanity matters.

## Control and Compensation

***We believe creators deserve to be in full control of our work, how it is used, and what we are paid for it.***

The development of generative AI algorithms should never come at the expense of the livelihood of the creators whose work has made the algorithms possible in the first place. Right now, generative AI companies are benefiting handsomely from algorithms they’ve “**trained**” on millions of pieces of our work that they haven’t paid a cent for, even though without the work of creators as input, these systems would not work at all. Creators have not been given a way to opt out of these **training corpuses** if we don’t want to participate, and proposed opt-out mechanisms would be burdensome at best, unworkable at worst, and ineffective with respect to works already “ingested” for AI use.

Additionally, the de facto impunity of AI companies with regard to the infringement of unregistered copyrights in Web content is an injustice, and a violation of international copyright treaties. Unless AI companies are required to obtain permission to use human-created works for AI “training” and development, they will pay only enough to mitigate their risk from litigation for copyright infringement, and even then will only pay those rights holders with pockets deep enough to sue them. It is currently not feasible in the US to register copyright in most Web content; and without the possibility of recovering attorneys’ fees, litigation for infringement of unregistered copyrights is effectively impossible. This is why there have been some lawsuits filed against generative AI developers for infringement of copyrights in books – which are easy to register, but not for infringement of (almost always unregistered) copyrights in Web content.

The Copyright Office has admitted the need for change in this arena, but has failed to act despite repeated calls for action by both the NWU and, among others, newspaper publishers. We maintain that this state of affairs is untenable, both from a moral rights perspective and in terms of US copyright law and fundamental fairness to Web creators.

## Transparency

***We believe that without real transparency, generative AI technology can’t operate ethically.***

One of the hallmarks of generative AI is the lack of transparency provided by the

corporations that sell their generative AI systems, or that provide generative AI on a “software-as-a-service” (SaaS) basis. They are unwilling (or, in some cases, unable, because of the way they have chosen to build their systems) to disclose the details of their algorithms — from the full corpus of works they’ve used to train the system, to how the algorithm weighs different inputs and picks outputs. In some cases, this is a true technical limitation; but in others, it’s a convenient excuse for companies to hide from having to compensate or credit the work that they’re benefitting from.

Generative AI companies must make changes to offer true transparency about their training data, the nature of their models and their output. This includes being open about the full list of content used to train each of these systems.

## Accountability

***We believe that users and providers of AI systems and services are responsible for their use.***

It is not JUST the users who have a duty to consider how their use of these systems impacts the human creators behind the training corpuses, as well as their own final audiences. It is also the developers' responsibility to ensure that their systems fairly compensate and credit the human creators whose work the systems rely on, and that the outputs of these systems meet high standards of accuracy and ethical integrity. All this is true regardless of who is deemed legally liable for the output of generative AI.

## Integrity

***We believe it is crucial to ensure that our audiences are not misled by the output of generative AI models.***

Generative AI systems do not “think” or “know” things. These systems work based on patterns found in large datasets. They cannot be anywhere, or see anything; they cannot interrogate sources or distinguish fact from fiction. Therefore, they cannot by definition be journalists, and to employ them as such is malpractice. As such the outputs that are generated by this technology will tend to replicate whatever common falsehoods, myths, and misunderstandings show up in the training corpus. This doesn’t include only lies and erroneous “facts,” but also unconscious biases and conscious bigotries. As a result, generative AI technology can often produce incorrect, misleading, or otherwise harmful information, passing it off as fact.

In the realm of journalism, where the credibility of factual claims is of critical importance, this is the least appropriate use case for generative AI. The dangers of this technological fact are far-reaching. AI can be used to generate voices or images that can be nearly impossible for people to distinguish as fabrications, or text that cites false or erroneously plagiarized information that end users might not know to verify, leading to a “GIGO” scenario (garbage in, garbage out). If left unchecked — e.g., if we don’t demand that developers incorporate guardrails and/or tools to alert end-users to the perils of trusting synthetic text/images/audio, the consequences will be dire.

With these foundational principles in mind, the following are our more specific policy requests:

## POLICY PLATFORM

### 1. Control

- (a) Creators should never be replaced by generative AI. If generative AI is used, it should be as a tool to assist human workers and augment our creative work, not as a replacement. It should be workers, not employers, who determine what is and is not assistive and augmentative. This technology should always supplement, never supplant creative work.
- (b) Creators should not be required by employers or clients to use generative AI in our work.
- (c) Employers must disclose to creators if any materials we are given have been generated in whole or in part by AI, or are based on AI-generated material.
- (d) Any use of human works for generative AI requires the permission of the creators of those works used as input. This must be done only on an “opt in” basis, either as an individual or as part of **collective licensing**.
- (e) “Opt in” systems should be easy to find, simple and quick to use, clear, and well advertised to creators.
- (f) Any “opt in” should be readily and effectively revocable at any time.
- (g) This should all remain true even if the human creator is not the copyright holder of the work. Too often, creators see our previous work used to enrich employers and companies in new ways, without our permission and without compensation.

### 2. Compensation

- (a) Creators should be compensated for all work used for AI at every stage, including, but not limited to:
  - i. Compensation for works already ingested in AI development.
  - ii. Compensation for future ingestion.
  - iii. Compensation for future use by generative AI systems and services,
  - iv. A share of revenues from generative AI **software-as-a-service (SaaS)** subscriptions and revenues from generative-AI outputs.

- (b) Any fair compensation strategy must ensure that creators are paid an appropriate rate for our work.

Once it is made clear in the law that AI companies may not use human-created works for AI “training” without permission, and once it is made clear in the law that creative workers are free to organize, market our work collectively, and bargain collectively with AI companies and other users of our work, it will be possible for AI companies and creators to discuss the permissible uses, price and terms for using our work. If they want to use our work, AI companies will need to offer sufficient payments to motivate creators to opt-in to a licensing scheme.

- (c) Use for AI development should not be deemed to fall under **fair use** or similar **exceptions to copyright**. Congress should explicitly exclude use for “training” of generative AI from the statutory definition of “fair use”.
- (d) Compensation for use in AI development should go to the human creators of the work, regardless of copyright ownership.
- (e) Compensation strategies might include, among other things, a **collective licensing** scheme.
- i. Creative workers must be freed to organize and market our works and rights collectively, without fear of possible antitrust sanctions, before we can organize collectively to decide the form of a licensing scheme and organization, what users and uses we may want to license, or the terms of those licenses.
  - ii. Any **reproduction rights organization (RRO)** carrying out collective licensing on behalf of creative workers should be a member-governed creator organization, on the model of e.g. worker co-ops or producer co-ops.

### 3. Credit, Labeling, and Transparency

- (a) Any work that was derived in whole or in part by generative AI should include credit (“attribution”) to the authors whose original creations were used to train the AI used to generate that work.
- (b) Credit entails a link to an online index that details all ingested work in the **corpus** (e.g. by **identifiers** such as an ISBN, URL, ISCC, etc.) used by the generative AI system(s) involved in the generation of the work.
- (c) Work generated in whole or in part by AI should be labeled to protect audiences from work that is misleading or incorrect. Labeling entails some kind of mark, as appropriate for the medium in question, that indicates to the readers, viewers, or listeners that something has been generated with the help of AI.
- (d) All AI-generated works should include **both** credit and labeling, not just one or the other. Together they serve two main purposes. The first is to give audiences



insight into how something was made. The second is to allow creators to see where their work has been used to generate content. Without both of these tools, we won't be able to assert our economic and moral rights.

## 4. Fair Contracts

- (a) AI development was not anticipated or paid for in past contracts (including work-for-hire contracts or contracts in which “all” rights were sold and licensed). As such, these contracts should not be read to allow for use of these works in AI training. Congress should clarify in the definition of works created for hire, that rights to use for AI development are retained by human creators, even with respect to works created for hire, unless those rights were *explicitly* assigned in writing.
- (b) Future contracts should explicitly ask creators for permission to use our work in AI systems. Those contracts should specify exactly which generative AI systems a work will be used for, and not simply be a blanket agreement for any and all uses.
- (c) Creators should be offered a percentage of revenue that is generated using our content. That may include revenue from subscriptions to AI services or work created using algorithms that ingested our material, as well as any revenue a publisher derives from licensing our content to AI companies.
- (d) Creators should have a say in the data governance of our clients and employers as well as that of platforms and distributors of our work. This way creators, if we choose, can ensure protections and controls on distribution and use of our work.
- (e) Freelancers and self-publishers should be afforded the right to bargain collectively with publishers, platforms/distributors, and AI companies without fear of violating **antitrust law**. An antitrust exemption for creators of intellectual property could be modeled on the longstanding antitrust exemption for agricultural cooperatives.

## 5. Compliance with the **Berne Convention** and other copyright treaties

- (a) US laws and regulations:
  - i. “**Fair use**”: For Congress (by statute) or the courts (through case law) to define copying and ingestion for AI development as “fair use” without consent or compensation would violate the Berne Convention **three step test** for permissible **exceptions and limitations to copyright**, and be wrong. Congress should explicitly clarify that such use is not permitted as “fair use.” Leaving the ambiguity in current “fair use” law to the courts to sort out would be unfair to creators who lack the deep pockets to match multi-billion dollar AI companies lawyer for lawyer, and would deny any meaningful redress to Web content creators who can't recover attorney's fees for infringement of works that are prohibitively costly and time consuming to register.

- ii. Registration formalities: The Berne Convention prohibits all “**formalities**,” and the WIPO Copyright Treaty requires that effective redress be available for all infringements of copyright. But the US requires timely registration of copyright as a prerequisite for recovery of attorneys’ fees or statutory damages, even if a lawsuit for infringement is successful. Congress should, by statute, direct the US Copyright Office to promulgate, by a date certain, a practical and affordable bulk registration procedure for Web content, including **granular and dynamic content**, which is currently prohibitively expensive and burdensome to register.
- iii. Congress must enact legislation explicitly protecting and providing effective redress for violations of authors’ **moral rights**, independent of copyright ownership or any mechanisms for enforcement of economic rights such as civil litigation by rightsholders or criminal prosecutions for copyright infringement.

(b) European Union laws and regulations:

As both US and EU creators have noted, AI companies have interpreted the exceptions to copyright for “Text and Data Mining” (TDM) in Articles 3 and 4 of the 2019 EU “Directive on Copyright in the Digital Single Market” as allowing ingestion of copyrighted works for AI training without permission or payment. This was not the intent of the Directive, and if allowed to stand, this interpretation of the Directive would violate the three-step test in the Berne Convention for permissible exceptions and limitations to copyright. The EU should amend the Directive or issue authoritative interpretative guidance that the TDM exception does not include use for AI “training.”

- (c) Other countries: All countries that are parties to the Berne Convention and other copyright treaties must ensure that their laws comply with these treaties. All proposed laws and exceptions for AI development in any country must be evaluated against the requirements of the Berne Convention and other treaties for protection of **authors’ rights**, including the three-step test for permissible exceptions and limitations, prohibition of formalities, and recognition and redress for violations of moral rights.

## Glossary

**antitrust law:** In the US, “antitrust” laws — which were intended to break up exploitative corporate monopolies — have sometimes been misapplied to organizations of workers. In response, Congress has enacted a limited exception to antitrust law for labor unions, but it applies only to unions of employees, and not to unions of “self-employed” freelancers and self-publishers. Another exception to antitrust law for co-ops applies to agricultural producers, but not to producers of intellectual property. The threat of antitrust enforcement chills organizing by freelance and self-published creative workers. There is case law that establishes, and we maintain, that freelance workers who compete for the same work as employees are covered under the labor exemption to the antitrust laws. But Congress could and should clarify and make this explicit, to remove the chilling effect on our organizing.

**authors:** In copyright law, “authors” include creators of both text and visual works (graphic artists, photographers, etc.) in all genres, media, and publication formats, not just writers or authors of books.

**authors’ rights:** In international copyright law and treaties, “authors’ rights” include both economic rights (referred to in the US as “copyright”) and moral rights (largely unimplemented in US law).

**Berne Convention:** The Berne Convention and other copyright treaties set a baseline for protection of both the economic and moral rights of authors. Parties to these treaties, including the US, can provide additional protections for creators, but must provide at least the minimum required by these treaties. There are good reasons why these minimal rights should be respected in national laws, independent of treaty obligations, but these treaties provide an important set of standards for national laws.

**collective licensing:** Collective licensing (also known as collective rights management) is a form of copyright licensing where creators form organizations (“reproduction rights organizations,” as defined below) that license rights to copyrighted materials *en masse* and pay creators from the fees they charge licensing organizations.

**corpus:** The complete data set — which may include text, images, audio, or other media — which an AI system analyzes and uses to generate output.

**exceptions and limitations to copyright:** The Berne Convention allows parties to the treaty to create “exceptions” and “limitations” to copyright in their national laws that authorize copying without permission of or payment to creators, but only if those exceptions and limitations are consistent with the “three-step test” (see below).

**fair use:** A legal defense under US law that allows someone to use a copyrighted work without permission or payment for certain limited purposes of comment, news reporting, criticism, teaching, scholarship or research. Many generative AI producing companies have argued that their “ingestion” (see definition below) of our work to develop their algorithms and generate derivative works from a corpus of our works should be considered fair use. Under current law, whether something is “fair use” can only be determined by the courts through years of expensive litigation, which favors those with the deepest pockets for a legal war of attrition.

**formalities:** In copyright law, “formalities” include any administrative, labeling, or other prerequisites for copyright protection, such as registration, fees, or inclusion of a copyright notice or symbol. The Berne Convention requires that protection of copyright be automatic, without any formalities.

**generative AI:** A blanket term used here to encompass a variety of machine learning models and applications that can generate text, images, audio, code, and other types of content. This includes systems like ChatGPT, Bing AI, Midjourney, Firefly, Stable Diffusion, Lex, Sudowrite, and more.

**granular and dynamic Web content:** Rather than storing and serving up “static” Web pages, most modern Web content management systems such as Wordpress store individual content elements (blocks of text, images, etc.) as separate files or in a database, and construct each Web page for each visitor on-the-fly as a customized assembly of multiple elements. The elements themselves can change each time an article or other item is updated. This makes each version of each content element a separate “work” for purposes of copyright registration. If the elements are small and numerous, it is prohibitively time-

consuming and the fees are prohibitive, especially for granular and dynamic text elements, to register copyright for each of them.

**identifiers:** Identifiers are metadata embedded in, attached to, or derived from a work that allow it to be identified uniquely and distinguished from other works. ISBNs are assigned to books. URLs are used to retrieve Web content. An ISCC is a “hash” derived from a file that can be used not only to identify the work but also to help assess whether different files contain versions of the same work.

**ingestion:** The process of taking a corpus (see definition above) and preparing and processing that data before it can be used for training (see definition below). This step includes breaking raw data down into a series of elements (tokens) that can then be further processed.

**just transition:** A framework of ideas and interventions that allow for a transition from one economic model, or way of doing things, without harming or displacing people.

**moral rights:** Moral rights are a category of rights that are considered to be human rights of creators regarding their creative work, independent of ownership of copyright. Moral rights guaranteed by the Berne Convention include:

- The right of attribution (i.e. the right to be properly credited as the author/creator of a work).
- The right to the integrity of the work (i.e. the right to object to alteration, distortion, or mutilation of the work that would harm the author’s reputation).

Moral rights are independent of economic rights. Even if a creator assigns their copyright to someone else, they maintain their moral rights. The United States has ratified the Berne Convention and thereby committed itself to implement protections for moral rights, but has not yet done so for authors of written works.

**reproduction rights organization (RRO):** Also known as collective management organizations (CMOs), RROs are organizations that manage the rights to creators’ work and make deals on their behalf with companies or organizations that might want to license that work. RROs come in a variety of forms and operate in many different ways based on different industries, laws, and national and regional regulations. Different RROs operate by legal statute, voluntary representation, and hybrids in between. Modern examples in the US include ASCAP, BMI, and SESAC, which enabled songwriters to get paid when broadcast radio came out.

**software as a service (SaaS):** SaaS is a software licensing and delivery model in which software is licensed on a subscription basis and is centrally hosted. SaaS is also known as on-demand software, web-based software, or web-hosted software. It includes any software that doesn’t need to be downloaded to a computer and runs on servers “in the cloud”, e.g.. Zoom, Slack, Google Docs, Microsoft Office 365, Dropbox, Salesforce. Most AI services are offered on a SaaS basis.

**three step test:** The Berne Convention allows exceptions or limitations to copyright only (1) “in certain special cases” and provided that they do not (2) “conflict with a normal exploitation of the work” or (3) “unreasonably prejudice the legitimate interests of the author.”

**training:** The process of analyzing a corpus of human-created content as input to language

models and algorithms used by AI systems to generate derivative output is often referred to anthropomorphically as “training.” We use that term here, as the term in common usage, while recognizing that AI systems have no human intelligence and are incapable of “learning.”

**work/works:** In copyright law, a “work” is an individual copyrighted item (text, image, audio, video, etc.). But our “works” in this legal sense are also our “work”: the fruits of our labor through the creation of which we earn our livelihoods.