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

212-254-0279
nwu@nwu.org
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 <u>and</u> 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 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.