The Authors Guild

**Authors Guild Policy Proposals Regarding
the Development and Use of Generative AI**

The Authors Guild believes that it is crucial for our culture and the future of democracy to ensure that our literature and arts remain vibrant and diverse. Generative artificial intelligence technologies can develop new content from huge volumes of existing material are about to have a significant impact on human creators and the future of our arts. It is imperative that we approach their development and use with a full understanding they will have on the writing profession and the arts more broadly—and with respect for human creators and copyright.

To protect the future of journalism, literature, and the arts, we must develop sensible policies and regulations governing the development and use of generative AI. Three specific issues are paramount, and the need to deal with them is urgent.

Issue 1: Generative AI Has Been Trained on Human-Authored Works to Mimic those Works Without Consent, Compensation, or Credit

To train their systems to generate new works, developers of artificial intelligences like GPT have copied millions of copyrighted works from the internet or illegally compiled databases without permission, relying on fair use under copyright law to do so. The works are copied multiple times in the course of training, and the output often closely resembles the original copied material.

This is a very different scenario from prior court cases such as *Authors Guild v. Google* where the use claimed by Google was to create a database to make books searchable (and only snippets readable). Here, almost the entire corpus of human creative work that is available online has been scanned to create works that compete in the market for those copied. It is grand theft in the extreme and should not be permitted. Yet, there is no certainty that courts will determine that copying works for AI training is not fair use. This is because many recent fair use cases have adopted an exaggerated version of the test laid out by the Supreme Court *in Campbell v. Acuff Rose* that favors finding fair use for background copying where the output is non-infringing. As such, we are asking Congress to help prevent evisceration of the creative professions before this egregious, massive taking becomes de facto uncompensated and uncredited.

It is simply unfair that generative AI systems have been created from existing works of human authorship and arts without compensation. Generative AI would not exist but for the works of human creators. It is worth noting that in some cases, generative AI developers have privately licensed some specific training material, an indication of its value. But most writers and other creators have been left out in the cold, their work expropriated without limit. These creators should be paid for the use of their work to train AI systems.

Position and Proposal: AI developers be required to obtain licenses for the copyrighted works that they use to train their AI; this can be achieved though collective licensing

a. *Collective Licenses Need to be Established*

AI developers cannot be expected to expend great effort to obtain licenses from individual copyright holders. Large-scale licensing schemes will need to be created so that AI developers can license the rights they need for AI training purposes in bulk from a single or several entities. Collective licensing is an established concept and an appropriate one. Collective licenses, some mandated by statute, are already available in the U.S. for certain uses of musical compositions, sound recordings and television programming; and they are widely used in Europe and much of the rest of the world for most creative industries, including books and journalism. They are an effective means of paying creators and publishers where licensing on a one-off basis creates market inefficiencies. For instance, in much of Europe, libraries and universities pay into collective licenses for photocopying and other specified uses. For many years now, the Authors Registry and the Authors Coalition of America have distributed royalties received from foreign collective licenses to U.S. authors.

Licensing human creation for AI training will not solve all the issues that AI will present to writers and other creative professionals, but it will put some money back into the pockets of creators and their distributors and at least partially compensate them for their efforts, so that many might be able to remain in the creative professions. It is a step in the right direction towards respecting human creativity.

How Collective Licensing Would Work

Collective licensing organizations could be established, or existing ones could be expanded, to offer bulk licenses for certain kinds of works (e.g., text, images, music) to developers to allow those works to be used for training generative AI systems. Each collecting society would then distribute the fees to participating copyright owners. What works are licensed and who qualifies to receive distributions, as well as formulas for who gets what percentage, will be worked out by each entity and its members. For example, a collective license could be created for the use of a database of books to train AI. The copyright owners who participate would share in the revenue collected from developers who use books to train AI. A mechanism would need to be developed to determine how to apportion payments. This is doable and is already done for many other uses throughout the world.

Required Legislation: Legislation may be necessary to avoid U.S. antitrust law violations arising from collective negotiations and agreement on terms. Legislation also would be necessary to adopt a provision for the Copyright Royalty Board (CRB) to step in and conduct a proceeding to determine rates and terms should the copyright owners and AI developers fail to reach an agreement by a certain date. The Copyright Act contains several provisions that permit negotiations between owners and certain users of a certain class of works, "notwithstanding any provisions of the antitrust laws." If the parties fail to reach a voluntary agreement, these provisions provide that the CRB will conduct proceedings to determine rates and terms. See, e.g., §§ 17 U.S.C. Secs. 114, 116, 118). Aside from the antitrust concerns, legislation is not necessary to form a collective licensing organization that enters into voluntary licensing arrangements, but having the CRB determine rates if the parties fail to agree could be helpful in encouraging the parties to come to the table – provided the CRB sets the rates in a fair manner.

b. Extended Collective Licensing

Another option that the U.S. government might consider is legislation authorizing extended collective licensing (ECL) for AI uses for certain works. Under an ECL, the government authorizes a collective management organization ("CMO") to negotiate licenses for a particular class of works for a particular type of use. The licenses they negotiate apply to CMO members and non-members alike, but non-members have the ability to opt out. The ECL framework is authorized by legislation that specify the categories of works (e.g., text or images) and uses (e.g., training generative AI) that the license would apply to.

ECLs could be administered by the U.S. Copyright Office. A CMO representing copyright owners of works in an authorized category would be permitted to seek authorization from the Register of Copyrights to issue licenses on behalf of both members and non-members of the CMO. The CMO would be required to show that it represents a broad group of impacted rightsholders, that its membership consents to an ECL, and that it adheres to sufficient standards of transparency, accountability, and good governance. Once authorized, a CMO would be entitled to negotiate royalty rates and terms with AI developers.

In 2011, the Copyright Office looked at the potential for ECLs for mass digitization and issued a Notice of Intent to obtain public comment on the proposal.[1] The Office concluded that it was premature to create ECLs in 2017 due to a lack of interest among stakeholders (in turn due to a lack of understanding of the coming technologies). However, copyright owners are interested in such solutions today as their works are already being used with impunity to train generative AI to produce works that compete with theirs.

Required Legislation: Legislation would be necessary to authorize ECLs and to permit an opt-out regime rather than opt-in, as well as to avoid antitrust issues that might arise from collective licenses.

c. Collective licensing for AI output that incorporates pre-existing human-authored works

Many AI systems can be prompted to produce work similar to other works or in the style of a certain writer or artist. These outputs, while clearly taken from a particular human creator, may not rise to copyright infringement under current U.S. copyright law, which requires that the infringing works be "substantially similar" to the original work. When these outputs are sold in the marketplace in competition with an author's or artist's own work, however, they harm the market for the original work, amounting to an uncompensated taking of the author or artist's expression and raising issues of authenticity and unfair competition. Right of publicity and unfair competition laws can assist in these cases but will not always apply to or redress this kind of unfair taking. Where human-authored works are used to train AI are incorporated in any given output, the human creators should also be compensated for the use. We understand that technologies that would enable tracking the use of data from input to a particular output are not yet fully developed but are feasible.

---

[1] https://www.copyright.gov/policy/massdigitization/

Required Legislation: Legislation that requires permission or payment for substantial uses of works in AI outputs within a few years would incentivize the development and implementation of such technologies, and would enable copyright owners to be paid for the use of their works in AI generated outputs.

Issue 2: Removal of Copyright Information in the Ingest/Training Process

When ingesting copyrighted works to train AI or to create databases for training AI, information about the work, author and owner is stripped out. This means that when an author or artist's work is incorporated in a generative AI output, they are not credited for the work, any more than they are paid for it, even when the output is in the style of the artist or closely resembles their work. This is equally unfair.

Proposal: We recommend amending section 1202 of the Copyright Act so that it is a violation to intentionally remove "copyright management information" from a copyrighted work without permission of the copyright owner, whether or not it can be proven that the removal was knowingly done to induce or enable infringement.

Section 1202 currently prohibits the removal or alteration of copyright management information (defined as information such as the title, author, owner, or terms of use for a work with which it is associated), but only if the removal is done "knowing[ly], or… [with] reasonable grounds to know, that it will induce, enable, facilitate, or conceal an infringement…" Because the AI developers who have used copyrighted works to train AI without authorization have generally claimed fair use, "knowledge" can be difficult to prove. Even if the copyright owner asserts that the use was not a fair use, developers will argue that they did not know that the removal of metadata would induce infringement because they believed that the use was non-infringing.

Required Legislation: Section 1202(b) could be amended to delete the last phrase starting with "knowing, or… having reasonable grounds to know, that it will induce…"

Alternatively, a new section could be added that states:

"No person shall, without the authority of the copyright owner or the law, intentionally remove or alter any copyright management information in the course of using a work to train artificial intelligence."

Issue 3: Market Dilution Due to AI Generated Works Competing with Human Works

Generative AI can produce works exponentially faster and cheaper than the human authored works they are based on; and humans won't be able to compete with the volume of AI-generated works that flood the market.

Position: AI-generated works should not receive copyright protection

Pursuant to U.S. case law and Copyright Office policy, the copyright law has long been understood to protect only human authorship. This means that works generated using AI systems

should not receive copyright protection unless and to the extent there's also demonstrable human expression in the work; and only the human authored portions of work created using AI should receive copyright protection.

Nevertheless, some argue that works generated solely by AI should be copyrightable (and patentable) because the Copyright Act does not specify that authorship must be human. For instance, an individual named Stephen Thaler sued in the District Court for the District of Columbia to appeal the Copyright Office's refusal to register a visual work that he claimed was generated solely by AI. He brought a similar case against the USPTO for denying a claim of inventorship by AI, and he has just filed a petition for a writ of certiorari with the Supreme Court in that case.

There are several reasons that AI-generated works should not receive copyright protection, but one of them is that it if AI-generated works were entitled to the same protection as human-created works, it would give them an artificial leverage in the marketplace and inevitably crowd and dilute the marketplace to the point that copyright incentives no longer function as intended. Few human creators will be able to earn enough to sustain a profession and the human quality of work produced by professionals – those who have talent and have trained in their careers for many years – will disappear.

Copyright law provides incentives to humans to create and publish original works. AI systems do not need incentives to generate new works; nor are AI-generated works original in the sense of "original authorship" required under the Copyright Act. They are merely derivative of the works the AI was trained on, and lack any new meaning or expression, unlike human created collages and other derivative works. Copyright protects original works of authorship, and authorship as defined by the Supreme Court is a product of human intellect. Humans necessarily put some of themselves, their thoughts, emotions, experience, and personalities into the works they create; and an original work of authorship must contain that spark of human intellect. AI technologies as known today are not capable of adding such sparks of creativity. They can mimic human creativity, but only by regurgitating what they have been trained on. They do nothing to promote the "Progress of the arts and sciences…" – the very basis of copyright law under the Constitution.

Required Legislation: No legislation is required now. If, however, courts find that AI authorship is copyrightable, legislation will be required to clarify that Congress did not intend for non-human authorship to be included in section 102 of the Copyright Act.

In sum:

Generative artificial intelligence raises significant issues for human writers, artists, musicians, and other creators—and the public that enjoys the fruits of their labors. These issues must begin to be addressed now and continued to be monitored in the future so that our knowledge, arts, and culture can continue to grow and thrive. The Authors Guild stands ready to provide our expertise in this vital aspect of our cultural heritage.