7.15.2023

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 human-created literature and arts remain vibrant and diverse. Generative artificial intelligence—computer programming that can develop new content from huge volumes of existing material—is about to have a significant impact on human creators and the future of our arts. It is imperative that we approach this issue with a full understanding of the impact AI technologies of today and tomorrow 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 Uses Human-Authored Works to Mimic those Works Without Consent, Compensation, and Credit

Developers of artificial intelligences like GPT have copied millions of copyrighted works from the internet or illegally compiled databases without permission, relying on claims of fair use under U.S. copyright law to do so. The works are not only copied many times in the course of compiling the databases and the continuous training but are embedded in the very fabric of the AI programs.

This is a very different scenario from that presented in 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, many millions of copyrighted works—almost the entire corpus of human creative works that are available online, including from some pirate websites—have been copied to create material that competes in the market with the copied works. These generative AI technologies would not exist if not for this taking; indeed, the copyrighted creative works are part of the fabric of the workings of generative AI, and they are intended to be used to mimic and regurgitate the language, stories, style, and expression copied. It is grand theft in the extreme and should not be permitted without the permission of the copyright holder.

Reproducing copyrighted works to "train" generative AI should not be considered "fair use." However, some recent fair use cases have adopted an exaggerated version of the test laid out by the Supreme Court in *Campbell v. Acuff Rose* (510 U.S. 569, 1994) that favors finding fair use for background copying where the use is deemed "transformative," and the output is non-infringing. While the Supreme Court's recent decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith* (598 U.S. ___, 2023) reminds lower courts to consider the degree of transformation against the commercialism of the use, as well as the other factors, the fair use doctrine is still ill-suited for mass use cases such as here where damage to the value of each

particular work might be hard to establish. As such, we are asking Congress to help prevent the evisceration of the creative professions before this egregious, massive taking becomes de facto, uncompensated, and uncredited.

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. It is only fair that these creators and those who have invested in them be paid for the use of their work in AI systems.

Proposal: AI developers should obtain permission for the use of copyrighted works in generative AI; this can be achieved though collective licensing in the marketplace

Free Market Collective Licenses Need to Be Established Where Necessary

It is not efficient to require AI developers to track down hundreds of thousands if not millions of individual creators to obtain licenses, and they have demonstrated an unwillingness to do so. Large-scale licensing schemes will need to be created so that AI developers can license the rights of individual creators in bulk from a single or several entities. Collective licensing is an established concept and an appropriate one for marketplaces where individual creators (rather than corporate copyright owners) often retain the copyright, and so are numerous and may be hard to locate and obtain license from. We recommend establishing free market collective licenses rather than statutory ones where rates are set by the government.

Free market based collective licenses are already available in the U.S. by voluntary agreement in many forms in the fields of musical compositions and sound recordings, and they are widely used in Europe and much of the rest of the world for most creative industries, including books and journalism. Collective licenses are an effective means of paying creators and publishers where obtaining a license from every individual creator and publisher creates market inefficiencies. For instance, in much of Europe, libraries and universities pay into collective licenses for photocopying and other specified uses. In the U.S., the Artists Rights Society licenses artwork for various uses, and for many years now, the Authors Registry, the Authors Coalition of America, and more recently the American Society for Collective Rights Licensing (ASCRL) 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 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) to AI developers to allow those works to be used for training generative AI systems. Each collecting organization would collect and 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, can be worked out by the collecting organization 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 already done for many other uses throughout the world.

Proposed Legislation: ***Legislation may be necessary to avoid U.S. antitrust law violations arising from collective negotiations and agreement on terms on behalf of members.*** Without some kind of recognized exception from the antitrust laws, both the rightsholders and users who enter into negotiations collectively could be exposed to antitrust lawsuits. Most collective licenses currently set out in the Copyright Act that allow for directly negotiations include a proviso "notwithstanding the antitrust laws." (See, e.g., §§ 17 U.S.C. Secs. 112, 114, 118.)

Aside from the antitrust issues, legislation is not necessary to form an opt-in, voluntary collective licensing organization that negotiates and enters into licensing arrangements on behalf of its members.

Legislation could also provide for arbitration, or for Copyright Royalty Board (CRB) or another entity to mediate rates and terms, should the copyright owners and AI developers fail to reach an agreement by a certain date or on behalf of parties who were not party to negotiations for an agreement for the voluntary collective license each year.[1]

Extended Collective Licensing

For the mass, indiscriminate training of AI that has already taken place, where the AI companies cannot necessarily even identity all works that the AI was trained on, they need to obtain ex post facto blanket licenses that would cover all of the works—and pay for them. Extended collective licensing (ECL) would assist with this. It is a type of collective rights licensing where qualifying collective management organizations (CMOs) can negotiate licenses in the marketplace for a

[1]The Copyright Act contains several provisions that permit negotiations between owners and users of a certain class of works, "notwithstanding any provision of the antitrust laws," and if the parties fail to reach a voluntary agreement, for the CRB is to conduct proceedings to determine rates and terms. See, e.g., §§ 17 U.S.C. Secs. 116, 118, 119. Other countries use various forms of mediation to assist when there is a failure to reach agreement. See, e.g., Denmark's Consolidated Act on Copyright 2014 (Consolidated Act No. 1144 of October 23rd, 2014), Section 52 ("(1) In the absence of any result of negotiations on the making of agreements as mentioned in section 13(1), section 14, section 16 b, section 17(4), section 24 a and section 30 a, each party may demand mediation. (2) Demands for mediation shall be addressed to the Minister for Culture. The request may be made if one of the parties has broken off the negotiations or rejected a request for negotiations, or if the negotiations do not appear to lead to any result. (3) The mediation shall be made by a mediator to be appointed by the Minister for Culture. The mediation negotiations shall be based on the parties' proposal for a solution, if any. The mediator may propose to the parties to have the dispute settled by arbitration and may participate in the appointment of arbitrators. (4) The mediator may make proposals for the solution of the dispute and may demand that such a proposal be submitted to the competent bodies of the parties for adoption or rejection within a time-limit fixed by the mediator. The mediator shall notify the Minister for Culture of the outcome of the mediation..."), located at **https://www.wipo.int/wipolex/en/text/546839.**

specific type of use on behalf of a specific class of copyright owners, whether or not they are existing members of the organization, but there must be an effective mechanism for non-members to opt out of the licenses. Legislation is necessary to authorize these types of licenses (which otherwise would have to be on an opt-in basis) and are intended for mass use, where users cannot negotiate directly with all individual copyright holders due to their sheer numbers. This is particularly appropriate for uses where many of the rightsowners are individual creators and they are numerous. The ECL legislation authorizes collective management organizations that meet certain criteria (and an agency may be appointed to formally approve the organizations) to negotiate blanket licenses on behalf of the entire class on an opt-out basis.[2]

ECLs for AI training could be subject to authorization by the U.S. Copyright Office. 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 on behalf of the class.

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.[3] The Office concluded that it was premature to create ECLs in 2017 due to a lack of interest among certain stakeholders who did not see a need at the time. New AI technologies have changed that perspective, however, and many individual copyright owners are interested in such solutions today—as their works are already being used with impunity to train generative AI to produce material that competes with human creation.

Proposed Legislation: ***Legislation would be necessary to authorize ECLs and to permit an opt-out regime rather than opt-in. The ECLs should be authorized to provide licenses for use of particular classes of works in generative AI on an industry-by-industry basis, where the right holders are particularly numerous, such as where individual creators hold the right and, as such, it is particularly inefficient to provide licenses on an opt-in basis (i.e., where each rights holder must be identified, contacted, and permission obtained on an individual basis). Such an ECL could apply to past use only provided that the law is clear that permission needs to be obtained going forward.***

The Copyright Office would be given the authority to authorize organizations who represent a substantial number of the rightsholders of a particular class of works (e.g., literary works) to enter into agreements with the users—the generative AI companies—for the extended collective license. Rightsholders in those classes would have the right and ability to opt out of such licenses. The Copyright Office would issue regulations to ensure that robust notice was provided

[2] For examples of ECL's, see Swedish Copyright Act, Chapter 3a: Lag (1960:729), English translation located at https://www.wipo.int/wipolex/en/text/532409; Denmark's Consolidate Act on Copyright 2014 (Consolidated Act No. 1144 of October 23rd, 2014), Section 50-52 , English translation **located at https://www.wipo.int/wipolex/en/text/546839**; DIRECTIVE (EU) 2019/790 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, Article 12; **located at https://eur-lex.europa.eu/eli/dir/2019/790/oj**.

[3] **https://www.copyright.gov/policy/massdigitization/**

to the covered class of authors of their right to opt out and that the procedures for doing so are simple and readily available.

Issue 2: Generative AI Has Been Trained on Human-Authored Works to Mimic those Works Without Consent, Compensation, and Credit

Many generative AI systems can be prompted to produce outputs similar to other works or in the style of a certain author or artist or to allow a particular author's or artist's works to be incorporated into outputs. 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 expression in an infringing work be "substantially similar" to that of 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's or artist's expression and raising issues of authenticity and unfair competition. The 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.

Proposal: AI developers should require permission and pay compensation for "in the style of" works

Where human-authored works are incorporated in any given AI output, permission should be obtained from the human creator and compensation paid, if desired. 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. It should also be feasible to track and pay creators when their names or the titles of their works are used in prompts.

Proposed Legislation: ***Legislation that requires permission for substantial uses of works in AI outputs would incentivize the development and implementation of such technologies, and it would enable copyright owners to be paid for the use of their works in AI generated outputs.***

This could be framed as **an additional exclusive right under copyright or as a *sui generis* right.**

A federal right of publicity would also assist in protecting creators' rights to use of their identity in AI outputs.

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; humans won't be able to compete with the volume of AI-generated works that flood the market.

Proposal: AI-generated works should be clearly labeled

A labelling-requirement for AI generated material—with enforcement provisions that give it teeth—would protect consumers from being misled into purchasing or consuming AI-generated

content that they assumed was human-created. This would help reduce incentives to dump large quantities of low-quality AI generated content into online and other marketplaces. It would also protect consumers against fake imagery and videos that are passed off as authentic news, furthering Congress' goal of preventing AI-generated disinformation and scams from proliferating in media and markets.

Proposed Legislation: ***Legislation could require authors, publishers, platforms, and marketplaces to label AI-generated works and otherwise identify when a significant portion of content has been generated by AI. The legislation would also provide for private causes of action and public enforcement, and could be made part of an omnibus AI law.***

Proposal: 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 material generated using AI should not receive copyright protection unless and to the extent there's also sufficient, demonstrable original human expression in the work, and any AI-generated elements or portions of the work should be excluded from 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, 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 filed a petition for a writ of certiorari with the Supreme Court in that case.

If AI-generated works were entitled to the same protection as human-created works, it would incentivize the use of AI to generate content that mimics human-authored works in place of hiring human creators, and it would give AI outputs artificial leverage in the marketplace, inevitably crowding and diluting 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.

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. Humans necessarily put some of themselves, their thoughts, emotions, experience, and personalities into the works they create; 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.

Proposed 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.***

Proposal: AI companies should be required to disclose training data

AI companies should be required to publicly disclose training data of their models to ensure safety of the models, and prevent use of sensitive, harmful, or illegally harvested data in the training. This requirement will also further encourage AI developers to work with copyright owners to license works for AI uses, instead of relying on datasets created with pirated copies of the works, as has been the case so far.

Proposed Legislation: ***Legislation would be necessary to require training data disclosures, and can be made a part of an omnibus AI law.***

In sum:

Generative artificial intelligence raises significant issues for human writers, artists, and other creators—and the public that enjoys the fruits of their labors. These issues must begin to be addressed now and continue 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.