Response from Creative Commons

Submitted by Kat Walsh, General Counsel, on behalf of Creative Commons

Creative Commons is writing to address the questions posed in the Notice of Inquiry on Artificial Intelligence and Copyright, Docket 2023-6. This response addresses questions 1, 2, 4, 5, 8-10, 15, 18-22, and 32.

General Questions

(1,2):
Creative Commons is a nonprofit organization that helps overcome legal obstacles to the sharing of knowledge and creativity. Toward that end, we publish a suite of licenses and legal tools designed to give every person and organization in the world a free, simple, and standardized way to grant copyright permissions for creative and academic works; ensure proper attribution; and allow others to copy, distribute, and make use of those works. In addition, we work closely with major institutions and governments, including the United States government and the governments of numerous U.S. states, to create, adopt and implement open licensing and ensure the correct use of CC licenses and CC-licensed content.

A large portion of CC license users are artists and creators, particularly independent artists. These creators are choosing to give away some of the exclusive rights granted to them by default by applying CC licenses to their works, to best fulfill their own economic and creative purposes. We believe that generative AI presents both benefits and risks, but that there is tremendous potential for both the creative and scholarly communities.

Creative Commons has always sought out ways to harness new technology to serve the public interest and to support better sharing of creative content — sharing that is inclusive, just, equitable, reciprocal and sustainable. We support ways for creators to share their works as broadly and openly as they want, so that people can enjoy them globally without unnecessary barriers. We also advocate for policies that ensure new and existing creators are able to build on a shared commons, while respecting creators' legitimate interests in control and compensation for their creative expressions.

A founding insight of Creative Commons is that all creativity builds on the past. Creators necessarily learn from and train their own skills by engaging pre-existing works and artists — for instance, noticing the style in which musicians arrange notes, or building on surrealist styles initiated by visual artists. Star Wars invented the character of Luke Skywalker, but it built on the idea of the hero's journey, among many other elements from past works. People observe the ideas, styles, genres, and other tropes of past creativity, and use what they learn to create anew. No creativity happens in a vacuum, purely original and separate from what's come before.

Generative AI can function in a similar way. Just as people learn from past works, generative AI is trained on previous works, analyzing past materials in order to extract underlying ideas and

other information in order to build new works. Image generation tools develop representations of what images are supposed to look like by examining pre-existing works, associating terms like "dog" or "table" with shapes and colors such that a text prompt of those terms can then output images.

Given how digital technologies function, training AI in this way necessarily involves making temporary copies of images in order to analyze them. We think this sort of copying can and should be permissible under copyright law. There are certainly nuances when it comes to copyright's interaction with these tools, some of which we will address further in this document. But treating copying to train AI as per se infringing copyright would in effect shrink the commons and impede others' creativity in an over-broad way. It would expand copyright to give certain creators a monopoly over ideas, genres, and other concepts not limited to a specific creative expression, as well as over new tools for creativity.

We have concerns about this technology and its responsible development and use: the impact these tools will have on artists and creators' jobs and compensation, use of these tools to develop harmful misinformation, to exploit people's privacy (eg, their biometric data), or perpetuate biases, as well as the sustainability of the information commons. These issues particularly impact the open sharing community, which often is producing information, scholarship, and creative output in the public interest, without the same incentives for creation as many other content creators. But copyright is not the only lens to address these problems, and our response in general cautions against using copyright as a tool to address all social concerns in place of more appropriate remedies.

Toward these ends, we have had several consultations with our communities on AI policy and approach with the aim of growing and sustaining the commons of freely-available information, some of which will be referenced below. We have heard from many different people, including artists and other rightsholders as well as developers of AI technologies, and we believe that it is possible for policy to guide the development of AI in ways that benefit all. These conversations inform our responses, here.

(4):
Creative Commons has been working with the EU to refine its approach to AI regulation, the AI Act. While that law is not perfect, its overall approach to AI regulation is logical insofar as it focuses on harmful uses rather than the technology itself. That said, this approach may have difficulty addressing advances in the technology because of the broad scope of potential uses that it holds; any approach to regulating AI must be flexible and able to develop with changes in the technology. We are also concerned about the scope and content of some of the requirements on disclosure of training data, which, while generally desirable, may be implemented with requirements that are too burdensome for practical compliance.

We have also seen Japan's approach making permission to train AI explicit: in July 2021, Japan's Ministry of Economy, Trade, and Industry published its report on AI Governance in Japan which provisionally disavowed any need to institute "legally-binding horizontal

requirements for AI systems," while also stating that future discussions should "be implemented in consideration of not only risks but also potential benefits."[1] Keiko Nagaoka (Minister of Education, Culture, Sports, Science, and Technology) confirmed that existing Japanese law does not outlaw uses of copyrighted material for the training of generative AI. This policy clears up uncertainty and places Japan in a key position to be a center for AI development.

Additionally, CC believes that international consistency would be strongly desirable for AI regulation. Not only will this help individuals so they do not have to navigate a maze of regulations when trying to understand their rights in an individual jurisdiction, it can help to avoid a race to the bottom, where AI development concentrates in the place with the least regulation.

(5):
We don't believe new copyright legislation is warranted to address copyright-related issues with generative AI. In particular, we caution against the creation of new rights that would cover uses that currently do not require a license under limitations and exceptions to copyright.

We are concerned about new restrictions on the ability to make use of a copyrighted work in order to extract data from it to be processed by a machine. Treating copying to train AI as per se infringing copyright would in effect shrink the commons and impede others' creativity in an over-broad way. It would expand copyright to give certain creators a monopoly over ideas, genres, and other concepts not limited to a specific creative expression, as well as over new tools for creativity.

We also see the copyrightability of outputs as being adequately handled by existing copyright law: that works not authored by humans should not be eligible for copyright protection, and that works authored by humans should be eligible. There is a tremendous variety of tools and processes lumped together under "generative AI" which involve varying degrees of human involvement in the creative process, some of which the Copyright Office has already considered. We consider the approach of granting copyright to the human-authored parts of a work while exempting the AI-generated parts to be the correct approach, even where drawing the distinction in practice may be difficult. This problem is not unique to AI: authors have always created work building upon and remixing uncopyrightable and other public domain works, by remixing old songs, translating or editing old texts, incorporating government-authored works, and other such artistic components from the public domain.

(8, 8.1, 8.5):
We believe that, in general, use of copyrighted works for training should be fair use.

In particular, we believe the uses being made of these works are transformative, fitting in with a long line of cases that have found that non-consumptive, technological uses of creative works in ways that are unrelated to the expressive content of those works are transformative fair uses. In *Author's Guild v. Google*, the Second Circuit Court of Appeals found that Google's act of digitizing and storing copies of thousands of print books to create a text searchable database

---

[1] https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf.

was fair use, as Google's purpose was different from the purpose of the original authors. The books were like pieces of data that were necessary to build Google's book database, and they were not being used for their expressive content. It was necessary to use the copyrighted material to build a digital tool that permitted new ways of using print books that would be impossible otherwise; the books as part of Google's database served a very different purpose from their original purpose, which supported the finding of fair use in this case.

It is also similar to the search engine case *Kelly v. Arriba Soft*. In this case, search engine Arriba Soft copied and displayed copies of photographs as thumbnails to users. The court held that this use served a different and transformative purpose from the original purpose because Arriba Soft only copied Kelly's photographs to enable its search engine to function, and not because of their aesthetic value. The images here served a function as data for the tool, not as works of art to be enjoyed as such.

The works being used as data to train these AI systems function similarly. Copyrighted works are used as sources of data from which data is collected and used to form mathematical models that learn from this data. These models, which have drawn inferences from the works but do not contain exact copies, can then generate new works of computer authorship which are also not copies of the original works.

Andy Warhol Foundation v. Goldsmith also continues to support this conclusion that training generative AI is transformative fair use. *Warhol* directs lower courts to compare the specific ways that an original work and a secondary work are used. In that case, the Court held that two commercial magazine uses of a photo had identical purposes, suggesting that the fair use analysis of this factor would be different if the uses had differed in character and purpose. With generative AI, the use of copyrighted works for AI training has a fundamentally different purpose from the aesthetic purposes of those works. This use bears more resemblance to the transformative use in search engines than the similar uses of artistic display of the subject of the image in commercially published works.

Notably, training generative AI is separate from using generative AI. While it may be possible to generate works to use generative AI in ways that infringe upon pre-existing works and that compete in the market for the works that the models are trained upon, the training itself does not interfere with the original intended aesthetic use or value of the originals. Accordingly, the market considerations should be considered only where the work is being used in a form that duplicates the original–simply gleaning data from a work is not sufficient, as the newly created work is not a comparable market substitute. As *Warhol* noted, the "central question" in fair use is whether a secondary use "supersedes the objects" of the original work. Using creative works to learn about what makes them what they are does not interfere with or supplant those works.

(9):
Creative Commons recognizes the value of giving creators ways to define their preferences for how others use their works; we have a suite of six separate licenses that millions of creators use

to share their works with the commons according to their intentions for the work, built atop copyright law rather than replacing it.

We see value in approaches allowing rightsholders to affirmatively express preferences to opt in or opt out of training, particularly as addressing these preferences is likely to create more trust in the systems where the development has respected them. But we are skeptical of this as a new requirement in copyright law–instead, we would like to see voluntary schemes, similar to approaches to web scraping, which allow for standardized expression of these preferences without creating strict barriers to usage in cases where it may be appropriate–for example, use of this data in training AI models for academic research purposes.

For another example, CC is supportive of traditional knowledge labeling such as the scheme developed by Local Contexts.:[2] This scheme does not rely solely on rights in copyright law; instead, it calls for a voluntary standard as to how this material should be treated because copyright law is unsuited to the needs of these communities.

In our community consultations on this issue, we saw general support for this area of exploration but no clear consensus.[3] While many artists and content creators want clearer ways to signal their preferences for use of their works to train generative AI, their preferences vary. Between the poles of "all" and "nothing," there were gradations based on how generative AI was used specifically. For instance, they varied based on whether generative AI is used

- to edit a new creative work (similar to the way one might use Photoshop or another editing program to alter an image)
- to create content in the same category of the works it was trained on (i.e., using pictures to generate new pictures)
- to mimic a particular person or replace their work generally, or
- to mimic a particular person and replace their work to commercially pass themselves off as the artist (as opposed to doing a non-commercial homage, or a parody).

Views also varied based on who created and used the AI — whether done by researchers, nonprofits, or companies, for instance.

Many technology developers and users of AI systems also shared interest in defining better ways to respect creators' wishes. Put simply, if they could get a clear signal of the creators' intent with respect to AI training, then they would readily follow it. While they expressed concerns about over-broad requirements, the issue was not all-or-nothing.

While there was broad interest in better preference signals, there was no clear consensus on how to put them into practice, with concerns about social effects being expressed on both the pro and con sides of introducing it. The effect on the information commons is not yet clear; a hard requirement enshrined in law at this stage may lead to choices that do not serve the public interest. We are aware of numerous efforts from AI developers on systems that respect

---

[2] https://localcontexts.org/labels/traditional-knowledge-labels/
[3] https://creativecommons.org/2023/08/31/exploring-preference-signals-for-ai-training/

voluntary preference signals, and we would like to see further work and experimentation with voluntary schemes before supporting a legal requirement.

(10, 10.4):

CC is aware that Creative Commons licensing is often pointed to as an example of extending creator choice through licensing schemes, and as the steward of those licenses we caution against relying on that example. Primarily, the CC licenses are drafted explicitly to extend only as far as copyright, and not to override relevant limitations and exceptions; it does not create contractual obligations where a use would be allowable under fair use. That is, where a use of a copyrighted work would be permitted by fair use, the CC license does not impose its conditions, such as attribution requirements or commercial use restrictions.

We are skeptical of a collective licensing approach because of the importance of maintaining limitations and exceptions to copyright for de minimis and transformative uses.

Transparency & Recordkeeping

(15):

CC generally supports measures that provide the public with transparency into AI models and the data used to train them. Transparency can help build trust in AI models by allowing others to "look under the hood" and investigate how those models work. However, these measures should not be so onerous as to make it impractical or impossible for developers to develop their technology. Moreover, transparency should not be enforced through existing or newly created copyright measures.

We believe disclosure of training data in the development of AI models is desirable for numerous reasons, including suitability for academic and research purposes, supporting the sustainability of communities and institutions where training data is created, and crediting sources. However, we do not support an obligation for disclosure of training data that is rooted merely in copyright.

For data sets of copyrighted works, we encourage collection and disclosure of licensing and authorship information as far as is possible when it is available for those works and would expect best efforts to be made to retain them. At the same time, we recognize there may be some important data sets or use cases where such information is not reasonably possible to collect or distribute.

Copyrightability

(18-21):

Human creativity should be eligible for copyright, even where combined with AI-generated work. The current standards for copyrightability of a creative output should remain as they are in the context of AI, where human authorship is eligible for copyright and machine authorship is not.

We recognize that determining which parts of a work are authored by a human in such situations will not always be clear–indeed, as determining authorship of individual parts of a work is not always clear in works created without AI assistance.

Copyright protection for the code should provide sufficient incentives to develop these systems without relying on additional rights in the outputs. Indeed, granting developers rights in the outputs would both stifle creative uses of these tools and be inconsistent with previous precedent. It is true that the technology shapes the artistic output of the tool–as does every tool for creating artwork. Tools for creating increasingly sophisticated artworks have been treated as just that, tools, and merely the implements that help creators bring forth their artistic visions. The Supreme Court looked at the relationship between artists and their mechanical tools vis-a-vis copyright over 100 years ago in *Burrow-Giles Lithographic Co. v. Sarony* and found that works produced with the help of artistic tools only qualify for protection insofar as the artists who use them add their own personal creativity to their outputs. We believe that works produced with the assistance of generative AI should be treated in a similar way.

Infringement

(22):
It is possible for AI-generated works to implicate the exclusive rights of existing rightsholders in some cases. We cannot enumerate all cases specifically; however, there are many possible ways this can happen, and we urge caution and flexibility in determining when an output has infringed exclusive rights. It is possible for an output to be infringing; it is also possible that an output including material under copyright may be using it in a way that is a fair use.

Liability for any infringement in this case should depend heavily on the facts of the situation, as the nature of generative AI tools and usage of those tools can vary greatly. For example, where a user has made a knowing attempt to get copyrighted material they have previously had access to as output from an AI tool (by directing prompts specifically toward that output, for example), that user has simply used the tool to reproduce the material, in the same way that one might use other tools to create a copy of an existing work.

We would caution against considering a simple attempt to make reference to a rightsholder's work as an infringement—for example, a user requesting artwork in the style of an existing artist. Copyright law already contemplates such cases, and already deals with some ambiguity that depends on facts: was a creator simply being inspired by the style of an artist, which is not an infringement and is critical to free expression and the development of artistic genres, or attempting to reproduce copyrightable elements in an infringing manner?

The debate about how copyright should apply to generative AI has often been cast in all-or-nothing terms—does it infringe on pre-existing copyrights or not? The answer to this question is certainly that generative AI *can* infringe on other works, but just as easily it may not. The tools that already exist in copyright, like the tests for substantial similarity and fair use, are

sufficient to address these questions. Our concern is that overregulation of generative AI could easily destroy this powerful tool that enables new ways for people to express their creative visions and share their works with the world.

Other questions

(32):
CC agrees with long-standing doctrine that style is not protected by copyright. As the Supreme Court recently recognized in *Andy Warhol Foundation v. Goldsmith*, the idea-expression doctrine provides a safeguard against copyright encroaching on First Amendment rights. If copyright begins to recognize protection over style because of generative AI, it runs the risk of granting rights to ideas in ways that are contrary to the fundamental principles of copyright law and that can harm the ability of future creators to express their artistic visions.

Conclusion

These questions about how copyright law should apply to generative AI are often difficult, fact-specific, and touch on areas other than copyright as means to solution. Copyright is an appealing road to look down for numerous reasons, particularly as it comes with easily available remedies and contractual alterations, but many of the questions presented imply more difficult questions about labor, ethics, misinformation, and other areas which the Copyright Office is not the correct place to address. As such, we suggest caution and flexibility in addressing generative AI. As with many new technologies, it presents great promise along with its negative effects, and we are concerned about addressing those effects with requirements that do not adequately consider the tradeoffs.

Creative Commons has seen its licenses and legal tools used as the legal infrastructure for numerous websites and millions upon millions of creative works by rightsholders who see the value of open sharing to a more creative, more knowledgeable, more equitable, and more innovative society. We have advocated for copyright as a balance between the individual rightsholder and the public interest. Limitations and exceptions to copyright form an essential part of US copyright law in contemplation of that public interest, and in this response we ask that those limitations and exceptions be considered as the essential feature that they are rather than simply considering how best to assign the exclusive rights.