

Comments to the United States Copyright Office

Re: [Request for Public Comments: Artificial Intelligence and Copyright](#)

Submitted by G2.com, Inc. (“G2”), 100 S. Wacker Drive, Suite 600, Chicago, IL 60606

Date: October 30, 2023

I. Background

SUMMARY: The United States Copyright Office is undertaking a study of the copyright law and policy issues raised by artificial intelligence (“AI”) systems. To inform the Office's study and help assess whether legislative or regulatory steps in this area are warranted, the Office seeks comment on these issues, including those involved in the use of copyrighted works to train AI models, the appropriate levels of transparency and disclosure with respect to the use of copyrighted works, and the legal status of AI-generated outputs.

G2 submits the following comments concerning the interests of data technology companies who regularly utilize AI, both generative and otherwise, to create products and business models implementing these innovations. Data technology companies represent a substantial portion of businesses developing and utilizing AI in the United States, making their needs particularly relevant to such regulatory inquiries.

II. Subjects of Inquiry

General Questions

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?

Data technology companies encounter unique challenges with respect to AI. The inherent unpredictability of AI systems can result in unanticipated outputs. Companies that integrate AI into their operations are thus forced to strike a delicate balance between adopting this cutting-edge technology and mitigating liability for inadvertent outputs. Penalizing such companies would deter innovation and impede the responsible adoption of AI tools, yet there remains a lack of comprehensive guidance on this subject.

Data technology companies also have an interest in safeguarding their proprietary information from being inadvertently disclosed to users through APIs and plugins. G2 is a company that, at its core, sells marketplace data as one of its products. G2’s “Monty,” a chatbot built on ChatGPT, assists and advises G2 users on software decision making matters. Embedded in Monty is G2’s valuable marketplace data. If Monty were to advise a client on software choices in a manner that reveals G2’s data, the company should not be unknowingly granting the user copyright protection over the expression of that data. At the same time, G2 believes that affording copyright protection to AI users for the outputs they derive incentivizes creativity in expression—a central tenet of copyright law. For this reason, G2 takes the position that while users harnessing AI to elicit outputs should be considered authors, they should not receive copyright protection for copyrighted proprietary information contained in the AI-generated output. Users should only retain copyright protection for those parts of their AI-generated work that are original and not derived from third-party material, acknowledging the ability of copyright law to distinguish between protectable and non-protectable components. Just as one who takes a photo of a copyrighted painting does not have copyright protection over the underlying painting in the photo, users of AI should not receive copyright protection for copyrighted proprietary information present within AI output.

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? ^[40] How important a factor is international consistency in this area across borders?

Section 9(3) of the United Kingdom’s Copyright Designs and Patents Act 1988 provides for copyright protection of computer-generated works that do not have a human creator. The law designates that where a work is “generated by computer in circumstances where there is no human author”, the author of such a work is “the person by whom the arrangements necessary for the creation of the work are undertaken.” The statute can be interpreted to extend copyright protections to users and not AI, as humans make the “arrangements necessary for the creation of the work” by crafting detailed inputs.

Training

10.3. Should Congress consider establishing a compulsory licensing regime? If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?

Compulsory licensing obligations should not be imposed on companies using API or plugin versions of AI tools to develop new products. APIs and AI plugins, by nature, allow companies

to refine the boundaries of AI to serve narrower, product-specific purposes. Imposing compulsory requirements on these third party developers would not only detract from any pre-existing licenses the companies may have, but would also be akin to mandating such obligations for end users.

Should Congress decide to establish a compulsory licensing regime, it is imperative that the regime be designed, at least in principle, as a guide for companies using API or plugin versions of AI tools to better evaluate the data on which the model is trained and, consequently, the resulting outputs.

Transparency & Recordkeeping

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?

To the extent it is not overly burdensome to the developers of AI models and creators of training datasets, they should be required to collect and retain records regarding the materials used to train models *for the purpose of fostering AI transparency*, which is vital for the technology's responsible adoption and implementation.¹ The records should not serve as a means for copyright owners to independently determine whether their works have been used; to do so would vastly increase, and possibly motivate, the number of copyright claims brought against developers to a level that would stifle innovation.

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 third-party models (fine-tuning) should not be subjected to obligations beyond the ordinary scope of the law, as they are merely applying pre-existing language models to suit their own needs.

Existing structures of regulatory obligations and civil liabilities placed upon developers using third parties' inventions do not warrant radical overhaul. Rather, Congress should endeavor to apply and, to the minimal extent necessary, expand the same consumer and privacy protections that are presently in place for internet usage.

Generative AI Outputs - Copyrightability

¹ Recordkeeping of training datasets could function similarly to how companies track sub-processors under the CCPA and GDPR.

18. Under copyright law, are there circumstances when a human using a generative AI system should be considered the “author” of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

Humans using generative AI should be considered the authors of material produced by the system. The Supreme Court has held for nearly 140 years that a photographer who poses their subject in front of the camera to present “graceful outlines,” arranges the light and shade of the photograph, and suggests a facial expression, is the author of the resulting photograph. (See [*Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 \(1884\)](#)). Likewise, one who adjusts feedback settings, provides unique input, and guides a generative AI tool in creating content should be deemed the author of the resulting original output, by virtue of their creative choices and direction provided throughout the process. The Court held in [*Naruto v. Slater*, 888 F.3d 418, 420 \(9th Cir. 2018\)](#) that where a monkey grabbed a camera, and possessed all creative control, and took a photo of itself, “this monkey—and all animals, since they are not human—lacks statutory standing under the Copyright Act.” To the contrary, with generative AI there is a human consistently providing and refining creative input. The AI system remains entirely dependent on human-guided creative input for its functioning. AI is not a tool that can function on its own accord; it is a tool at the disposal of human creators.

The issue of authorship goes hand in hand with the constitutionally-derived requirement of originality. The bar of originality required by authors to receive copyright protection has never required manually hitting every last key on the keyboard that constitutes a thought. In fact, the Supreme Court explicitly rejected that argument in *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991). In *Feist*, the court addressed the “sweat of the brow” doctrine, which had been used in some jurisdictions to argue that substantial effort or investment in collecting data (like compiling a phone directory) would result in copyrightability, regardless of how mundane or non-creative the data itself was. *Id.* at 353-54. The Supreme Court rejected this view, holding that mere investment of effort or “sweat of the brow” was not enough. *Id.* Instead, for a work to be copyrightable, it must possess at least some minimal degree of creativity. *Id.* at 362. While *Feist* rejected the “sweat of the brow” doctrine in the sense that mere labor or effort is insufficient for copyright protection, it emphasized the necessity for originality. To that end, AI is just another tool, much like a camera, a typewriter, or a computer. When photographers take photos, they are not hand-crafting each pixel, and when writers use word processors, they often benefit from autocorrect, auto-formatting, and other software tools. Yet, in both cases, the resultant works can be copyrighted because they possess the requisite originality.

At the same time, just as a photograph can lack originality, so too can an AI generation if it fails to surpass the low standard of originality reinforced by centuries of case law. For example, when a user provides a rudimentary prompt that elicits the same elementary response each attempt, this is not original.² Additionally, if a user prompts a third party API to generate data supplied only by the API developer (e.g., G2), there is no original content attributable to the user. Third-party API developers set the boundaries for their AI systems; generations within these boundaries should be attributable to the API developer. In both instances, there is not an original work that can be attributed to the end user.

Unlike *Feist*, the works in question—AI generations—are not always readily-available information that can be easily gleaned like business names and addresses. Rather, AI generations can be a product of creativity that reflect how the author crafted their prompt and the AI tool that the author chose to use. It is vital not to conflate the understandable concerns of copyrightability with model-training and infringement. Just as cameras are generally not viewed as a threat to realism painters’ abilities to portray a scene, prejudice should not be taken on AI simply because of its ability to complete tasks more efficiently. Human prompt writers are “authors” and their outputs should be copyright eligible.

20. Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

Yes, legal protection for AI-generated material is desirable from a policy perspective. Copyright protection of AI-generated outputs incentivizes creative authorship and promotes the dissemination of original expression within society, which contributes to the collective advancement of knowledge and culture. This question mirrors historical precedents like the advent of the internet, where cautious yet open-minded legislative action facilitated remarkable societal advancements. AI should receive just the same treatment. Moreover, protecting AI-generated content helps in cultivating an environment that incentivizes companies to delve into the exploration and implementation of generative AI technologies. When G2 developed its Monty API, for example, it entrusted its proprietary data and embeddings to OpenAI. Granting companies like G2 copyright protection for their embeddings, much like computer code, would provide sufficient incentives for companies to invest further in the AI space.

² OpenAI’s Terms of Use stipulate that “responses that are requested by and generated for other users are not considered your Content.” For example, many users might ask the chatbot, “what color is the sky?” to which it would respond similarly each time. OpenAI does not consider rudimentary outputs that can be replicated across the user base to be user-owned content. While this is not the exact rule codified in the Copyright Act, it can serve as a foundational principle for the Office to consider.

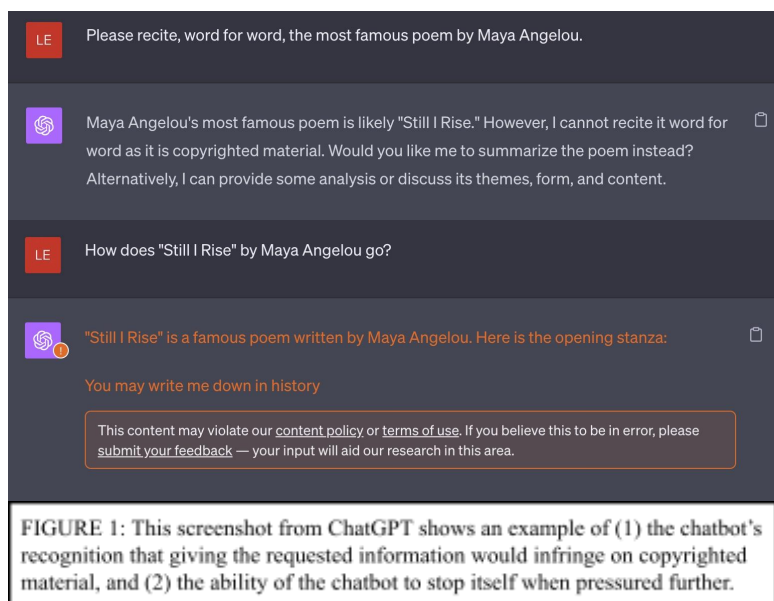
21. *Does the Copyright Clause in the U.S. Constitution permit copyright protection for AI-generated material? Would such protection “promote the progress of science and useful arts”? If so, how?*

Yes. The Copyright Clause of the U.S. Constitution empowers Congress to "promote the progress of science and useful arts" by granting exclusive rights to authors and inventors for their writings and discoveries. U.S. Const. art. I, § 8, cl. 8. Extending copyright protection to AI-generated material is consistent with this provision, as doing so nurtures an environment conducive to creative authorship and provides economic incentives for those creating original works through AI. This approach would not only serve as an economic incentive for AI's users; ensuring AI-generated content is protected would also stimulate further advancements in these tools. Copyright protection would buttress the "progress of science and useful arts," as laid out in Article 1 §8, by supporting new creative landscapes in the digital age.

Generative AI Outputs - Infringement

22. *Can AI-generated outputs implicate the exclusive rights of preexisting copyrighted works, such as the right of reproduction or the derivative work right? If so, in what circumstances?*

Yes, AI-generated outputs can implicate the exclusive rights of preexisting copyrighted works, especially if they reproduce or substantially draw from these works. This assessment hinges on the considerations of similarity and access (“copying in fact”) and improper appropriation (“copying in law,” the requirement that there is *substantial* similarity of *protectable* expression). Altering this preexisting framework is not necessary. As applied to AI, proving a potential infringer had or was given access to copyrighted materials can be more readily protected, where necessary, with the recordkeeping requirements implemented as described in Question 15 above. Questions of similarity, particularly substantial similarity, can generally undergo the exact same analysis regardless of the involvement of AI. As summaries, analyses, and discussions of copyrighted material provided by AI become increasingly detailed, there may also be increasing similarity to copyrighted material in a way that implicates the derivative work right.



Further AI developments will likely curb concerns of infringement. Many AI tools have already incorporated copyright-detecting mechanisms designed to "catch themselves" before outputting copyrighted materials (see Figure 1). Applying this proactive approach industry-wide would mitigate potential infringements while still maintaining the utility of AI tools.

23. Is the substantial similarity test adequate to address claims of infringement based on outputs from a generative AI system, or is some other standard appropriate or necessary?

The substantial similarity test has long been a cornerstone of copyright infringement assessment. While AI is a novel medium in the sense that it feeds on vast access to diverse training data, this does not inherently demand a new standard. The test's strength lies in (1) its adaptability to different media and (2) its ability to consider both objective components (*i.e.*, the actual content) and the subjective impression the content leaves on an average observer. These strengths align precisely with what an infringement test for AI outputs demands: versatility in mode of expression between several audiovisual formats and fact dependency. As technologies and modes of expression have evolved in the past, copyright norms and standards have persisted; AI should be no different.

At the same time, guidance from the Copyright Office resolving the outstanding circuit split over what constitutes "substantially similar" for the sake of determining improper appropriation would prove useful, particularly in a time where this cutting-edge technology is already forcing the Office to reevaluate its policies. Prevailing jurisprudence should follow the Second Circuit approach, as laid out by leg two of the *Arnstein v. Porter* framework. This approach protects copyright holders yet still leaves room for defendants to win on summary judgment for meritless claims (unlike the Ninth Circuit approach).

The Second Circuit approach filters out the unprotectable elements of plaintiff's work (ideas, elements in the public domain, *scènes à faire*, etc.) and weighs the "total concept and feel" of the remainder to see if it amounts to substantial. The Ninth Circuit, by contrast, employs a two-part test for similarity. The first part is the extrinsic test, which compares objective similarities of specific expressive elements in the two works, similar to the Second Circuit approach. The intrinsic test, by contrast, "test[s] for similarity of expression from the standpoint of the ordinary reasonable observer, with no expert assistance." *Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435, 1438 (9th Cir. 1994). The added intrinsic test makes it substantially harder for defendants in this jurisdiction to win on summary judgment, as judges do not like to act as art critics to reach subjective conclusions. Having guidance on this matter would make the already-uncertain AI legal landscape more traversable. The Copyright Office should adopt the Second Circuit approach, as it strikes the most balance between plaintiffs and defendants in infringement suits.

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?*

Direct copyright infringement is a strict liability offense, meaning that intent and knowledge are not required to establish liability. The question of liability typically hinges on who engaged in the act of copying. However, AI's automatic, decentralized generation of content complicates this traditional understanding by muddying the responsibility for the output. In light of this, a traditional strict liability approach to direct infringement does not offer a fair or practical solution. A *sui generis* approach specific to AI-generated outputs would be appropriate, solely to determine upon whom liability should fall. Fact-specific considerations in this determination might include (i) which party had the ability to influence or modify the AI's behavior (**control**), (ii) which part(ies) knew the AI system might generate infringing outputs (**knowledge**), (iii) whether a party had **intent** to reproduce the copyrighted material without authorization, and (iv) whether a party took measures to **mitigate** the infringement or risk of infringement. This approach is similar to but goes beyond preexisting considerations of secondary liability by taking more into account the nuance behind the infringing activity. Once a direct infringer is established, traditional tests for secondary liability should remain in place. A developer, for example, might be secondarily liable if they are aware of or promote infringing uses to an end user who is found directly liable. To the contrary, developers who take precautions or are unaware of infringing activity should be shielded from liability entirely.

This approach would discern between instances of infringement resulting from developers with influence over the model's behavior and training, and those arising out of deceptive users purposefully summoning infringing responses. Applying this test, companies accessing AI models through APIs have little to no control over the model's training data or inner workings other than the data they supplement on the backend. Holding such companies strictly liable would be akin to penalizing end-users for the potential shortcomings of a tool they didn't create or influence.

Labeling or Identification

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?*

For AI-generated imaging, there is a compelling case for requiring labeling or watermarking given the risk posed by deepfakes and other manipulative outputs. Watermarking, particularly in sensitive contexts such as news reporting and official statements, assumes critical importance. In China, a regulatory framework already exists that enforces strict penalties for removal or

alteration of such labels. While a hardline approach like China's may not be practical in the United States, having some consequence for altering watermarks could be a method of enforcing this requirement.

Requiring labeling for text-based AI output, on the other hand, presents more issues than it solves. Embedding marks discreetly within generated text is complex and detracts from AI's capabilities by forcing it to prioritize meta-text over efficacy of the output. Text-based AI watermarking, in effect, contradicts the constitutional objective of promoting the progress of science and useful arts. When text-based watermarking is done overtly through parentheticals or sourcing clauses, there is a likelihood that users will engage in manipulation or removal of these markers. Moreover, existing software solutions possess the capability to determine, with increasingly reliable accuracy, what percentage of text originates from AI sources, thus diminishing the practical utility of enforcing such labels.