Copyright Registration of AI-Generated Works Checklist

A Practical Guidance® Checklist by Kirk A. Sigmon, Banner Witcoff

Kirk A. Sigmon
Banner Witcoff

This checklist outlines key considerations that attorneys should review when advising whether and how to copyright artificial intelligence (AI) and machine learning (ML)-generated works in the United States. The checklist provides a framework for documentation of human involvement in the creative process of an AI-generated work and for the preparation of a copyright application. It focuses on collecting information useful for both the application and for responding to follow-up by the U.S. Copyright Office.

As a preliminary matter, applicants should exercise caution when trying to copyright works generated using AI or ML models. The U.S. Copyright Office (the Office) carefully scrutinizes such applications. Specifically, the Office has issued guidance stating that individuals using AI/ML technology to create a work may claim protection "for their own contributions to that work," but if "a work's traditional elements of authorship were produced by a machine, the work lacks human authorship and the Office will not register it." (Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, pp. 4, 5.) The Office thereby draws a line between work of an author's "own original mental conception, to which [the author] gave visible form" and creative works of a machine (including simple mechanical reproductions). Burrow-Giles Lithographic Co. v. Sarony, 111 U.S. 53, 60 (1884).

Documenting human involvement in the creation of an AI or ML-generated work is important because (1) the Office expects applicants to explicitly distinguish between human and AI contributions in copyright applications, and (2) the Office sometimes requests additional information from applicants when evaluating possible limitations on a copyright application involving AI-generated content.

For more information on generative AI, see Generative Artificial Intelligence (AI) Resource Kit.

For an overview of the copyright registration process, including how to draft and file a copyright application, see Registration of Copyrights.

Document the Nature of the AI

The training and capabilities of an AI model can have significant impact upon its ability to contribute—or not contribute—to a creative work. For example, if a model is rudimentary (e.g., designed to remove compression artifacts from existing images, designed to add makeup to a human face, or the like), then it might be fairly presumed to be less likely to provide creative output. As such, more human creativity might be implied in the resultant creative work. That said, if a model is highly sophisticated and trained based on previously published works, that model might be assumed to more readily provide what appears to be a creative work with relatively minimal human effort.

- **Record model(s) used.** If an existing model (e.g., a model downloaded from the internet) was used, collect information regarding the model such as:
  - When it was retrieved
  - Where it was retrieved from
  - A recorded version number
  - The date and/or time the model was used –and–
  - Other similar information

- **Document known model uses.** Some generative models (such as Stable Diffusion) can generate wholly new images, whereas some other models (such as those used as plugins in photo editing suites) are trained to improve and otherwise modify existing images. It is generally easier to argue that the latter are similar to conventional photo editing tools.
- **Document model training process.** If available, document how the model was trained. This can include:
  - Documenting the training data that was used, including information such as:
    - Where the data originated
    - Who owned the data –and–
    - The format of the data
  - Documenting the training process itself, such as:
    - Which algorithms were used –and–
    - Which loss functions were used
  - Documenting, where applicable, whether the model is designed to continually learn, such as where it might receive further training as part of a feedback loop during use

Example: Some freely available Stable Diffusion models accessible through enthusiast websites are quite sophisticated and are trained to emulate specific authors' work. It may be relatively difficult to copyright their output because those models require very little effort to produce output that appears quite creative and because the models are designed to, in effect, create permutations of another author's previously published work. In those circumstances, applicants should endeavor to document as much human creative labor as possible (and should expect an uphill battle). With that said, other models, while equally sophisticated, are designed to simply clean up and/or otherwise enhance existing works. Copyrighting the output of these models seems significantly easier in no small part because they are roughly analogous to an advanced photo filter.

Document the Scope of AI Contribution

Once the relevant model(s) are identified, it is critical to understand how those models were used during creation of a particular work. This establishes a baseline for later assessing the scope of human involvement.

- **Record prompts and user-controllable parameters.** Collect any prompts used to generate the relevant work.
  - In the case that a model is configured to "remember" past prompts or is otherwise configured to generate output based on a plurality of previous prompts, collect all such prompts.
  - Where applicable, collect and record other user-controllable parameters, such as the number of steps, the selection of samplers, and seed values.
- **Document other inputs.** In some circumstances, a model might be used to enhance and/or otherwise modify an existing creative work, such as an image or sound file.
  - For example, some Stable Diffusion implementations have an "img2img" function which receives, as input, both a prompt and an input image and then outputs an image based on the prompt and the input image.
  - As another example, some language learning models are capable of proofreading (and providing recommended edits to) input text, such as a draft of a book.
  - If a model was used to enhance or otherwise modify a previous work, collect examples of the un-modified and as-modified work.
- **Collect metadata and logs.** Some models provide metadata and/or logs relating to the process via which a creative work may have been output. While this information is rarely relevant to the question of whether a human was involved in creating the work, it can nonetheless be useful when, for example, showing how a human author used an AI tool over time (e.g., iteratively improved outputs over time).

Example: If an author merely provides a single and simplistic prompt (e.g., "show me a cat") and attempts to copyright that output (a picture of a cat), the Office's guidance suggests that the application is more likely to face pushback. The result might be different if the same user iteratively provided various prompts over time to modify the image (e.g., "now add a flower," "now make the image drawn in crayon," and so forth).

Document Human Creative Labor

Once the scope of model use is understood and relevant information about the model's use is collected, it is then extremely important to use that information to understand and document the scope of human creativity in generating the work. Where possible, applicants should endeavor to show that the work is "basically one of human authorship," with the AI model "merely being an assisting instrument." 300 Compendium II of Copyright Office Practices 313.2.

Put more simply, documentation should show that "some element of human creativity must have occurred." Urantia Found. v. Kristen Maaherra, 114 F.3d 955, 957–59 (9th Cir. 1997). Applicants should endeavor to take a very broad approach to collecting information at this stage, recognizing that creative labor can include a wide variety of activities.

- **Determine extent of prompt engineering.** If the author spent time crafting a prompt or otherwise modifying relevant parameters, document the process by which they modified and eventually selected a prompt with the correct requirements (i.e., the trial-and-error process of finding the right prompt). Keep in mind that this process might be iterative over time in at least two ways:
  - Some models are instructed iteratively and can base subsequent output on a series of previous inputs. For example, an author might instruct a model "draw a house," then "draw a car in front of the house," and so forth. Accordingly, prompt engineering (i.e., creative work on the part of an author) might be evidenced by an author providing a series of instructions over some period of time.
  - Authors might spend time making edits to a single prompt and accompanying parameters to generate a desired image. For example, an author might add or remove words to a large, wordy prompt and re-submit the prompt with each edit to evaluate the resulting output of a model and, over time, endeavor to reach a desired outcome. This process itself evinces creative effort on the part of the user insofar as they, through their prompt engineering, sought a desired creative output.
- **Determine inputs other than prompts.** If the input was something other than a prompt (e.g., if the author provided an image to be modified by the model), determine whether the human author made changes to this input. For instance, if an author created and/or made edits to an image (e.g., to add or remove some object) before providing it to a model for further processing, this is arguably a degree of creative effort on the part of the author.
- **Ascertain any modifications to output.** If the author made changes to the output after the fact (e.g., modification of an AI-generated image using photo editing software), determine the scope and nature of these changes. This output modification is arguably creative effort on part of an author.

Example: If an author spends time preparing a draft of a book before providing it to a language learning model for proofreading and then makes further edits to the resultant output of the model, the author arguably has provided significant creative labor, even where the model is quite sophisticated and perfectly capable of generating a book's worth of content by itself. The model in this case might be analogized to a proofreader/editor that is but one step in an overall creative process.

Draft Application

In its guidance, the Office encourages applicants of AI-generated works to use the Standard Application and "provide a brief statement in the 'Author Created' field that describes the authorship that was contributed by a human." (Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, p. 5) This is an area where careful drafting can prevent significant pushback from the Office.

- **Disclose any and all AI involvement.** Make known the use of AI without trying to downplay it.
  - Attempts to hide or otherwise tone down AI involvement could be perceived as fraudulent and risk the validity of the application and/or registration. See 17 U.S.C. § 411(b)(1)(A); Unicolors, Inc. v. H&M Hennes & Mauritz, L.P., 142 S. Ct. 941, 948 (2022).
  - Definitely err on the side of over disclosure, even where it might come across as pedantic.
- **Focus on human involvement in brief statement.** Focus directly on the human contributions to the work, recognizing that AI should be little more than a tool in the creative process.
  - The Office's own example of such a statement ("[s]election, coordination, and arrangement of [describe human-authored content] created by the author and [describe AI content] generated by artificial intelligence") does precisely that.
  - Avoid phrasing that incorrectly suggests passive behavior on the part of the human author (e.g., "[Author] used [AI model] to generate picture," without more). · Explicitly exclude AI-generated content. The Office's guidance instructs that "AI-generated content that is more than de minimis should be explicitly excluded from the application."
- **Explicitly exclude AI-generated content.** The Office's guidance instructs that "AI-generated content that is more than de minimis should be explicitly excluded from the application."
  - Use the "Limitation of the Claim" section and under the "Material Excluded" heading (and/or via the "Note to the Copyright Office" field). In particular, applicants should use this section to disclaim any AI-generated aspects of a work that are clearly based on previously published works.

- Given the tenor of the Office's guidance, if your work involved generative AI in any way, it may be wise to include some sort of disclaimer of some material, however minimal. Otherwise, the Office may reach out with questions and devise its own limitations.
- A good example of how the Office excludes AI-generated work from a copyright registration is in their letter relating to the partially AI-generated comic book Zarya of the Dawn. In that letter, the Office acknowledges human authorship of "the Work's text as well as the selection, coordination, and arrangement of the Work's written and visual elements" but concludes that the "images . . . that were generated by the Midjourney technology are not the product of human authorship."

- **Use the "Note to the Copyright Office" field liberally.** The Standard Application permits applicants to provide freeform comments, and this field should be used to extensively detail the involvement of AI and forestall potential questions from the Office. For example:
  - Include information collected about the nature of the AI, including which model(s) were used, how the model(s) were trained, and like information.
  - Provide extensive details regarding the use of the model, including relevant prompts, parameters, and any other input to the model. –and–
  - Explain in as much detail as possible the extent of human creativity involved, including any modifications to model input/output, trial-and-error, and other human creative effort.
- **Do not list AI as an author.** An author must be human. Naruto v. Slater, 888 F.3d 418, 426 (9th Cir. 2018). As such, listing an AI as an author is not only incorrect, but invites scrutiny by the Office.
- **Prepare for follow-up questions.** If an application indicates that AI was used in the process of generating a creative work, the Office might contact the applicant with questions regarding the use of the AI. This is, in part, why so much early data collection is recommended; it makes the process of answering these inquiries significantly easier and prevents encouraging the Office to exclude excessive content from the application.

For a discussion of the copyright registration process, see Registration of Copyrights. For a general discussion of copyright law, see Copyright Fundamentals.

Best Practices for Follow-Up Questions

If the Office follows up regarding an application, the best approach is to be honest and comprehensive. When answering questions, keep the following in mind:

- **Focus on the "Modicum of Creativity" standard.** In the United States, the "the requisite level of creativity" for a copyrightable work "is extremely low; even a slight amount will suffice." Feist Publications, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 345 (1991). As such, even if AI was 99.9% involved in the creation of a work, a human author is still entitled to a copyright in their (admittedly small) contribution.
- **Do not downplay AI involvement.** While some applicants might be tempted to downplay (or outright attempt to hide) the extent of AI involvement in a work, this approach can border on untruthfulness and could, at minimum, invite scrutiny by the Office. Instead, be candid about the extent of the AI used—otherwise, the Office could assume that almost all of the work was AI-generated.
- **Tie human action to creative labor.** Focus on human action that involves creative effort and, where possible, tie such efforts back to known copyright principles. For example, focus your answers on:
  - Demonstration of human creative labor in the selection and arrangement of particular elements in an image (via prompts or not)
  - Coordination of various steps for a desired outcome (e.g., creation of an input image using a camera, providing that input image to a model, receipt of a modified version of the image) –or–
  - Revision and remixing of other content (e.g., other content used to train the model).

Kirk A. Sigmon, Attorney, Banner Witcoff

Kirk's work in the U.S. and in Asia, tied with his experience with Fortune 500 companies and startups, provides him the know-how to counsel clients at all stages of invention, patent prosecution, intellectual property enforcement, and litigation.

Kirk began his legal career in Tokyo, and routinely works with U.S., Japanese, Korean, Chinese, and European intellectual property matters. Kirk's cases have involved a broad range of technologies, including computer networking, cellular communications, video gaming, virtual reality, machine learning/artificial intelligence, military weapons systems, blockchain technologies, aerospace flight systems, video encoding, petroleum engineering, optoelectronics, data storage, magnetics, agronomy, and toys. Kirk has been certified by the University of Hong Kong as a FinTech professional and is also an IBM-certified Machine Learning Professional, a Google-certified cybersecurity professional, and a Government Blockchain Association-certified Blockchain Legal Specialist. Kirk also speaks Japanese and is actively studying Korean.

One of *Managing IP's* Rising Stars, Kirk has successfully represented both plaintiffs and defendants in multimillion-dollar patent infringement trials in federal court. He has counseled Fortune 500 companies on topics including patent portfolio management and intellectual property enforcement.

An active member of the business community both in Washington, D.C., and abroad, Kirk has counseled startups and spoken at startup and M&A conferences around the world. Kirk is also a frequent contributor to the blog *Patent Arcade*, where he writes on video game intellectual property law.

Before joining Banner Witcoff, in addition to his intellectual property practice, Kirk worked with clients in response to Department of Justice, Consumer Financial Protection Bureau, and Securities and Exchange Commission investigations that involved technical issues such as data security.

Kirk devotes a significant amount of his time to charity and pro bono matters. He has successfully represented a Guantanamo Bay detainee in military proceedings, victims of sexual violence in U-Visa proceedings, and numerous adults and minors in other immigration proceedings. He also routinely prepares testamentary documents for indigent senior citizens. Outside of the law, Kirk is the head of a charity events organization in Washington, D.C., where he seeks to provide fundraising opportunity to local nonprofits and causes.

LexisNexis®