

**Before the
UNITED STATES COPYRIGHT OFFICE
Washington, D.C.**

In the Matter of:

Artificial Intelligence and Copyright

Docket No. 2023–6

COMMENTS OF SHUTTERSTOCK, INC.¹

Shutterstock, Inc. (“Shutterstock”) submits these Comments in response to the U.S. Copyright Office’s notice of inquiry and request for comments on topics related to generative artificial intelligence (“AI”) and copyright.

Shutterstock is a leading global creative platform connecting brands and businesses to quality content so they can create, build, or share their next big idea. Contributors upload their content to Shutterstock’s web properties in exchange for monetization of their content through royalty payments based on download activity. Shutterstock also owns certain content outright and makes it available to license as well. Shutterstock’s key offerings include a library of over 734,000,000 high-quality images, spanning “stock” and “editorial” photographs, vectors, and illustrations; video footage, cinema-grade video effects, and GIFs; music tracks and sound effects; and three-dimensional models that are used in a variety of industries such as advertising, media and video production, gaming, retail, education, design, and architecture. Content in Shutterstock’s web properties is captioned and sorted by particular subject matter and artistic style to facilitate easy searching. The vast majority of content contains metadata, which can be the generative power behind AI models.

Before Shutterstock content is licensed or downloaded, a preview or thumbnail can be viewed on Shutterstock’s website, but static images and videos are prominently watermarked to prevent them from being used without a license. Shutterstock content may be licensed via subscriptions that allow business customers to license a certain number of downloads of content in a given period, or through licenses for individual pieces of content. Shutterstock also provides tools to enable business and professional customers to best leverage the content which is licensed.

Shutterstock is at the forefront of issues related to generative AI. First, as an owner and licensor of content, Shutterstock has directly experienced the ways in which generative AI companies have developed products based on copyrighted content to which they do not necessarily have the rights. The fact that Shutterstock’s content is watermarked prior to being licensed has made it patently clear that certain AI companies have scraped Shutterstock’s content to train their AI models without consent or compensation to Shutterstock’s contributors. Second, Shutterstock has pioneered the process of providing ethically sourced content, enabling ethical output from AI

¹ These Comments are submitted on behalf of Shutterstock, Inc. by John Lapham, General Counsel.

models—AI models that respect and compensate copyright owners. This is demonstrated in one manner by Shutterstock’s AI Image Generator, which generates and customizes images while Shutterstock compensates creators for their data inputs.²

Shutterstock respectfully offers this submission in response to Questions 1, 5, 6, 6.1, 6.2, and 7.4 in the Copyright Office’s notice of inquiry. Specifically, Shutterstock wishes through this submission to make clear that ethically sourced AI content—i.e., output derived from AI models that are trained on content with permission and with appropriate credit and/or compensation to copyright owners—is feasible and currently available in the market. Shutterstock hopes that its experiences in this area will be helpful to the Copyright Office in shaping further AI policy.

I. Shutterstock’s Experience with Developing Ethical AI (Responsive to Questions 1, 6, 6.1, 6.2, and 7.4)

Shutterstock has played a foundational role in the development of ethical and responsible generative AI models, giving it unique insight into the effects that generative AI has on creators and copyright owners and the ways in which copyright-protected training materials may be appropriately used to train AI models.

Some time ago, Shutterstock became aware that certain AI companies were training their generative AI models on Shutterstock content without authorization. This became evident via third-party research into the contents of the datasets used to train prominent AI models.³ It was also evident because certain AI tools would generate images or videos with Shutterstock-like watermarks.⁴ Shutterstock had deep concern that its contributors’ copyrighted works were being copied for AI training datasets without authorization, transparency, or compensation.

Shutterstock thus set out to provide “ethical data sets” to companies driving machine-learning models. These generative AI models would be trained on ethically sourced datasets—that is, a dataset consisting of content used with copyright owners’ consent, the use of which would compensate creators for their inputs. One such result of this model (ethical data for ethical outputs) is Shutterstock’s AI Image Generator, developed based on a technological collaboration with OpenAI. This AI tool has been trained on images in the Shutterstock library and the rich metadata

² See Shutterstock AI, *AI Image Generator*, Shutterstock, <https://www.shutterstock.com/ai-image-generator> (last visited Oct. 30, 2023).

³ See, e.g., Andy Baio, *Exploring 12 Million of the 2.3 Billion Images Used to Train Stable Diffusion’s Image Generator*, Waxy (Aug. 30, 2022), <https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/>.

⁴ See, e.g., Benj Edwards, *AI-Generated Video of Will Smith Eating Spaghetti Astounds with Terrible Beauty*, Ars Technica (Mar. 30, 2023), <https://arstechnica.com/information-technology/2023/03/yes-virginia-there-is-ai-joy-in-seeing-fake-will-smith-ravenously-eat-spaghetti/>. However, and related to Question 7.4, watermarks both can be removed by an AI company before content is used in training data and are not always consistently reproduced in AI-generated output, which renders this a generally unreliable method to determine whether copyrighted content licensed by Shutterstock and other stock content companies was used in AI training data.

accompanying those images, such as image descriptions and keywords.⁵ Shutterstock has invested substantially in trust and safety, expanding on its existing safeguards to prevent the introduction of offensive content into the Shutterstock library and implementing guidelines to ensure that underrepresented groups are fairly depicted.⁶

Shutterstock has built robust demand for ethically sourced AI training data. It has partnered with multiple companies that are interested in training their AI models on licensed data from Shutterstock, including LG and Meta. Today, contributors may opt out of having their works included in licensed datasets, and contributors who do not opt out are compensated via a first-of-its-kind “Contributor Fund” that provides ongoing royalty payments connected to future data licensing.⁷

Shutterstock similarly already allows its contributors to opt out wholesale from the use of their content to train its AI Image Generator. For creators who do not opt out, Shutterstock also compensates them via the Contributor Fund. The Contributor Fund compensates creators when their content is used to train Shutterstock’s AI Image Generator, and further provides ongoing royalty payments connected to future licensing of generative AI output from the Image Generator.⁸ In turn, because Shutterstock’s AI Image Generator is built on licensed data, Shutterstock is confident in the rights clearances associated with any output—indeed, it was a leader in the industry and offered a similarly first-of-its-kind indemnification program for users.⁹

In addition to ensuring that training datasets used by Shutterstock for its own AI Image Generator and licensed to third parties are ethically sourced, and in addition to ensuring compensation for creators, Shutterstock has engaged in ongoing partnerships designed to develop best practices in responsible and ethical AI models. For example, Shutterstock is collaborating with AI for Good, an organization spearheaded by the United Nations’ International Telecommunication Union with a goal of establishing industry standards that recognize creators’ contributions to AI tools.¹⁰ Shutterstock is also mindful that responsible AI requires additional safeguards, such as ensuring the authenticity and provenance of content, and Shutterstock has been working with the Content

⁵ *Get to Know the AI-Generated Content Tool on Shutterstock*, Shutterstock Customer Support, https://support.shutterstock.com/s/article/Get-to-know-the-AI-generated-content-tool-on-Shutterstock?language=en_US (last visited Oct. 30, 2023).

⁶ *Id.*

⁷ *See Shutterstock Data Licensing and the Contributor Fund*, Shutterstock Contributor Support (last updated June 16, 2023), https://support.submit.shutterstock.com/s/article/Shutterstock-Data-Licensing-and-the-Contributor-Fund?language=en_US.

⁸ *See AI-Generated Content on Shutterstock: Contributor FAQ*, Shutterstock Contributor Support (last updated Sept. 8, 2023), https://support.submit.shutterstock.com/s/article/Shutterstock-ai-and-Computer-Vision-Contributor-FAQ?language=en_US.

⁹ *See* Bridget Johnston, *Introducing Indemnification for AI-Generated Images: An Industry First*, Shutterstock (July 11, 2023), <https://www.shutterstock.com/blog/ai-generated-images-indemnification>.

¹⁰ *See* Press Release, Shutterstock, Inc., Shutterstock and ITU’s AI for Good Collaborate to Advance Responsible AI (May 16, 2023), <https://www.shutterstock.com/press/20489>.

Authenticity Initiative to integrate technical standards and best practices into its AI- and non-AI creativity tools.¹¹

Shutterstock's experience shows that it is possible to have ethical generative AI, where creators have agency over whether their works are used for training AI models and receive compensation when that happens. Shutterstock's experience also shows that there is less legal and financial risk for the entire creative community if machine learning is based on these ethical sources, just as Shutterstock's commitment to ethical training data allowed it to commit to provide users of the Shutterstock AI Image Generator with additional assurance against future legal risk. The Copyright Office should shape policy that promotes this type of approach to ethical AI, while, where appropriate, discouraging the use of copyrighted content to train AI models without consent and/or compensation.

II. What the Copyright Office and Congress Can Do to Facilitate Ethical AI Models (Responsive to Questions 5 and 8)

As an initial step toward encouraging the ethical approaches to generative AI discussed above, the Copyright Office should provide guidance clarifying that the use of unlicensed copyrighted works in training data for AI models, and the reproduction and distribution of training data in AI model output, is not necessarily fair use by either AI developers or their users—particularly where there is substantially similar output and where the model competes with and replaces existing markets for licensed training data like the datasets Shutterstock provides. While the Copyright Act provides a framework for content owners and technology companies to navigate these issues, existing case law is unfortunately poorly situated to shed light on these issues, and so guidance from the Copyright Office is urgently needed.

In doing so, the Copyright Office can and should emphasize that this would not mean a death knell for generative AI. AI training regimes based on licensed content that fairly and transparently compensate creators—like those Shutterstock is pioneering—are not just possible, but are already succeeding, providing users with a level of trust in their engagement with AI models and helping to contribute to a safer environment of AI use.

Shutterstock greatly appreciates the Copyright Office's ongoing efforts to provide clarity in the area of generative AI. We would be happy to provide any further information or details to further inform your perspective.

Respectfully submitted,

/s/ John Lapham
John Lapham, General Counsel

¹¹ See Press Release, Shutterstock, Inc., Shutterstock Joins the Content Authenticity Initiative (July 25, 2023), <https://investor.shutterstock.com/news-releases/news-release-details/shutterstock-joins-content-authenticity-initiative>.

Shutterstock, Inc.
350 Fifth Avenue, 21st Floor
New York, NY 10118

October 30, 2023