gettyimages®

iStock™ Unsplash

195 Broadway, 10th Floor, New York, NY 10007 US
(800) IMAGERY (800) 462-4379 | www.gettyimages.comgetty

legal@gettyimages.com for related inquiries

October 30, 2023

Response to USCO Inquiry on Artificial Intelligence and Copyright

Docket No. 2023-6

Table of Contents

Executive Summary

Getty Images is a preeminent global visual content creator and marketplace. With customers from almost every country in the world and more than 830,000 customers in total, Getty Images works with businesses of all types and sizes to connect and compete in an increasingly digital and visual world. Getty Images offers a growing library of over 538 million visual assets (image and video) that delivers unmatched depth, breadth, and quality. Our library represents the work of more than 541,000 contributors. We are the partner of choice to major companies and organizations in the global editorial space and each year we cover more than 160,000 news, sport and entertainment events around the globe. We have partnered with trusted industry leader, NVIDIA, to develop "Generative AI by Getty Images" [1], a commercially safe, non-infringing, text-to-image AI generation tool that employs an AI Model trained exclusively on Getty Images content and data. We share the revenue generated by the commercialization of Generative AI by Getty Images with our world-class content creators, allowing them to continue to create more of the high-quality pre-shot imagery for which Getty Images is known and on which our customers depend, and we provide uncapped indemnification to authorized users of our tool for their publication and distribution of the images we deliver through it.

We believe that responsibly developed and properly licensed AI and Generative AI Models hold potential to provide significant benefits to consumers and creators alike; however, there are significant risks to consumers, creators, and the public interest when developers of AI Systems and AI Models exploit copyrighted content without permission from the relevant copyright holders.

The development of AI Models capable of generating visual images in response to text prompts requires access to high-quality visual works paired with high-quality metadata. There are numerous options for licensing large sets of high-quality visual works and associated metadata for use in connection with the training and development of AI Models, and responsible AI developers are licensing those datasets from providers such as Getty Images, Envato, Alamy and Shutterstock. Fully licensed, commercially safe, and non-infringing products, such as Generative AI by Getty Images and Firefly by Adobe, capable of generating high-quality images from text prompts are already available in the marketplace. Irresponsible developers, however, are electing to copy and exploit third parties' copyrighted images without permission in reliance on the argument (misguided in our view) that such unauthorized copying is defensible as a fair use, even when done at a large scale for commercial purposes, or they are simply hoping that their unauthorized use will not be detected and litigated by the relevant rightsholders. These unauthorized uses impede the growth of the licensing market for copyrighted works as training data, and the directly competitive offerings they facilitate aim squarely at the image licensing marketplace more generally to the detriment of the creators and copyright owners whose works were essential to the training process.

Getty Images (US), Inc. is a plaintiff in a lawsuit filed in the United States District Court for the District of Delaware against Stability AI, Inc. and Stability AI, Ltd. The case, Getty Images (US), Inc. v. Stability AI, Inc., Case No. 1:23-cv-00135-GBW, was originally filed on February 3, 2023, against Stability AI, Inc. Getty Images (US), Inc. filed an Amended Complaint on March 29, 2023, naming Stability AI, Ltd. as an

[1] See https://www.gettyimages.com/ai/generation/about

additional defendant. The dispute arises out of the defendants' alleged unauthorized reproduction of approximately 12 million images from Getty Images' websites, along with the accompanying captions and associated metadata, and use of the copied content in connection with Stability AI's Generative Artificial Intelligence model known as Stable Diffusion. Getty Images (US), Inc. has asserted claims for copyright infringement; removal, alteration and/or falsification of copyright management information; trademark infringement; unfair competition; trademark dilution; and deceptive trade practices. The Amended Complaint seeks, among other things, monetary damages and injunctive relief. For more information, please see the Amended Complaint attached as Appendix A hereto.

Getty Images advocates for the advancement of Generative AI technology in a manner that respects the rights of creators, sustains ongoing creation, and requires AI System and AI Model developers and providers to address misinformation, bias and misuse resulting from the use of their tools.

While this topic of the inquiry touches on many legal and policy issues, our responses below focus on Generative AI as AI Models and AI Systems that are trained with copyrighted images and video (as well as corresponding captions and metadata).

Definitions

We appreciate the USCO's inclusion of the Glossary of Key Terms and we have attempted to use them appropriately in our response. We would like to comment on two of the definitions:

***AI Model**: A combination of computer code and numerical values (or "weights," which is defined below) that is designed to accomplish a specified task. For example, an AI model may be designed to predict the next word or word fragment in a body of text. Examples of AI models include GPT-4, Stable Diffusion, and LLaMA.*

- It should be recognized that "AI Models" are built for different purposes using different techniques and not all AI Models are Generative AI models. However, we note that all of the examples referenced in the definition of "AI Model" are Generative AI Models.

- Non-generative AI Models exist that focus on tasks such as classification, recognition and prediction (non-Generative AI Models).[2] IBM Watson's predictive AI tool is an example of an AI Model that is not a Generative AI Model.

- While our responses focus on Generative AI Models and are phrased as such, we urge the USCO when reviewing other responses to be mindful that some respondents might be referring to AI Models in a broader sense and, therefore, their responses might not always apply equally to Generative AI Models.

***Generative AI**: An application of AI used to generate outputs in the form of expressive material such as text, images, audio, or video. Generative AI systems may take commands or instructions from a human*

[2] See https://medium.com/@nonsibiventures/ai-series-part-1-is-non-generative-ai-dead-c5559fbcde29

user, which are sometimes called "prompts." Examples of generative AI systems include Midjourney, OpenAI's ChatGPT, and Google's Bard.

- We note that the definition of "Generative AI" does not distinguish between Generative AI Models and Generative AI Systems, thus we have used "Generative AI Models" and "Generative AI Systems", separately and combined, as appropriate.

General Questions

1. *As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?*

Responsibly developed and properly licensed Generative AI Systems hold potential to provide significant benefits to consumers and creators alike. Allowing unregulated and unchecked development and deployment of Generative AI Systems and AI Models trained on copyrighted materials without permission from the copyright owners, however, creates the following risks:

For society: Absent regulation, society risks a flood of disinformation and synthetic content that will undermine the public's trust in institutions and each other. Without an obligation to identify generative content, the burden of verification will fall to the public, which is ill-equipped to distinguish between synthetic and/or manipulated content and authentic content. Worse, the debate over truth and authenticity will occur after generative content is published, if at all, giving it the opportunity to replicate and influence.

For the free press: Generative AI will impact existing business models of media organizations through the loss of referral traffic and links from social and social media. The harm to business models, together with the impact of AI Services offering a competing product built on existing intellectual property, will result in a corresponding reduction in investment in editorial coverage and creative content creation, further undermining the reporting of events and issues critical to an informed public and a functioning democracy.

For creatives: Copyright is an opt-in regime. We believe that developers of AI Models and AI Systems need to obtain affirmative consent from copyright holders prior to exploiting copyrighted works as training data because that exploitation typically involves numerous acts within the scope of the exclusive rights afforded under Section 106 of the Copyright Act. Opt-out schemes proposed by certain AI developers are neither consistent with U.S. copyright law nor sufficient to protect copyright holders' interests. Permitting Generative AI Models to be trained on copyrighted works without consent of the copyright holders enables the essentially limitless and immediate creation and distribution of directly competing content at low or no cost, which destroys incentives necessary for human creative endeavors. The development of Generative AI Systems and related services should not be at the expense of human creators essential for a vibrant and progressive society and the business partners who make their livelihoods possible. Generative AI has the

potential to have a major impact on the U.S. creative market, which contributes more than $1.016 trillion to the GDP and represents 4.9 million jobs or 3.2% of all American jobs.[3]

Fortunately, we know that it is not only possible to create responsible, non-infringing Generative AI Models and Generative AI Systems, but that there are already examples in the market today that produce commercially safe, high-quality outputs without infringing third-party rights.

In partnership with NVIDIA, Getty Images recently released "Generative AI by Getty Images", a Generative AI System that enables customers to generate commercially safe, high-quality synthetic visuals from a text prompt. The underlying AI Model was trained exclusively on licensed content and the output can be safely used for commercial purposes, without risk of infringing third-party rights. Importantly, and unlike with respect to text-to-image Generative AI Models developed outside of a licensed, permissions-based framework, copyright owners receive recurring compensation for the use of their content as training data for Generative AI by Getty Images. As Getty Images grows revenue from this Generative AI System, the associated rightsholders participate in any revenue realized from commercialization for their contributions to the tool. Other responsibly developed AI Systems and AI Models have also emerged, including Firefly by Adobe[4] and latent diffusion models developed and licensed by Bria,[5] which are trained on licensed content and are marketed as safe for commercial use.

Some AI Model developers have claimed that the requirement to comply with intellectual property laws and the need to license copyrighted works used as training data will hamper innovation. They point principally to their desire for immense quantities of content and the financial costs of obtaining licenses for such content as their rationale. As set out more specifically in response to Question 13 below, these claims are untrue. No differently than in any other area, innovation with respect to generative AI technologies can, and should, co-exist with respect for intellectual property laws.

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?*

Getty Images' editorial photography and video content is licensed by our customers to inform people, public policy, services, and markets, and we operate under an editorial standards policy to ensure that our coverage is free of obligation and has no conflict of interest in its creation or production. Our editorial offering is comprised of photos and videos covering more than 160,000 news, sport and entertainment events annually around the globe. We invest in a dedicated editorial team that includes over 115 staff photographers and videographers to generate our own coverage, in addition to coverage from our network of freelance photographers and videographers and content partners.

---

[3] See Facts and Figures on America's Creative Economy published by National Assembly of State Arts Agencies https://nasaa-arts.org/nasaa_research/facts-figures-on-americas-creative-economy

[4] See https://www.adobe.com/sensei/generative-ai/firefly.html

[5] See https://bria.ai/models/

Getty Images' creative content, comprised of images and videos, is released for commercial use and is licensed by our customers to produce impactful websites, digital media, social media, marketing campaigns, textbooks, movies, television and online video content. We distribute this content on behalf of a broad network of professional, semi-professional and amateur creators.

At Getty Images we offer our customers commercially safe AI Generated content through our proprietary Generative AI System, "Generative AI by Getty Images." This AI System was created in partnership with NVIDIA and trained exclusively on content we either own or have licensed from copyright holders who will participate in any revenue realized from commercialization.

Where Generative AI is developed and used responsibly, such as in "Generative AI by Getty Images", it can enhance creativity and enable inspiration in a commercially safe way, while opening new revenue streams for rightsholders whose works were used to develop the tool. In contrast, when Generative AI is developed using copyrighted content indiscriminately scraped from the internet without permission of the relevant rightsholders, including images of celebrities, newsworthy events, personal and confidential information, it can enable any actor to produce and distribute synthetic content at a scale that far exceeds our past experience. Such a flood of content, if unregulated and unchecked, can distort facts and leave the public with no basis to discern what is true and what is fiction.[6] Even absent malicious intent, many Generative AI applications produce factual errors and fictional information, in addition to propagating long-standing biases.

In addition, Generative AI Models that are unlawfully trained on proprietary content scraped from the internet without permission, including content belonging to Getty Images and its contributors, enable the essentially limitless and immediate creation and distribution of directly competing content at low or no cost without any consideration of, renumeration, or attribution to the original creators whose works were essential to the process. This naturally results in a significant negative impact to creators' livelihoods and their respective abilities to invest in new content.

3. *Please identify any papers or studies that you believe are relevant to this Notice. These may address, for example, the economic effects of generative AI on the creative industries or how different licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works in training AI models. The Office requests that commenters provide a hyperlink to the identified papers.*

No response.

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? How important a factor is international consistency in this area across borders?*

[6] See AI-generated deepfake image of pope Francis wearing a puffer jacket. https://www.theguardian.com/technology/2023/apr/23/pope-jacket-napalm-recipes-how-worrying-is-ai-rapid-growth or AI-generated image of an explosion near The Pentagon. https://apnews.com/article/pentagon-explosion-misinformation-stock-market-ai

We are highly supportive of efforts by governments and industry groups to create a set of consistent global standards applicable to the development and deployment of AI. It is essential that that those standards respect and protect the intellectual property rights of the owners of content on which Generative AI Models are trained. We also must ensure compliance with appropriate transparency standards that inform the public both about the content used to train the model and that outputs from the model are synthetically generated.

Regulatory harmonization may be difficult to accomplish but is important to the successful regulation of Generative AI in the United States and at an international level. We encourage U.S. regulators and legislators, as long-standing leaders in support of creative industries and intellectual property protection, to work with their global counterparts, including other members of the Berne Convention, to develop basic international norms and standards addressing consent of rightsholders to the use their content in training AI Models and basic transparency standards.

We view the most recent version of the draft European Union *AI Act*, as published by the European Parliament, an effective approach to addresses transparency standards in relation to both (i) source of training data that includes copyright works at the input stage;[7] and (ii) the labelling of outputs of Generative AI tools.[8] We believe that such transparency standards will help promote responsible innovation.

In February 2023, the United Kingdom, where there is an exemption to copyright law for text and data mining done when for non-commercial research purposes, withdrew potential amendments to that law that would have expanded text and data mining exemptions to include commercial purposes following collective outcry from the members of creative industries. The potentially devastating impact to the creative industries was acknowledged in a House of Lords Communications and Digital Committee report (issued January 2023)[9] that recommended the UK Intellectual Property Office pause its proposed changes to the text and data mining exemption immediately, finding that the proposed changes took insufficient account of the potential harm to creative industries. For the same reasons, we are opposed to interpretations of or changes to existing copyright laws in the U.S. that would permit unauthorized use of copyrighted works as training data for commercial applications of Generative AI Models.

---

[7] See European Parliament Proposed Article 28b(4), "Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system, shall in addition a) comply with the transparency obligations outlined in Article 52 (1), b) train, and where applicable, design and develop the foundation model in such a way as to ensure adequate safeguards against the generation of content in breach of Union law in line with the generally acknowledged state of the art, and without prejudice to fundamental rights, including the freedom of expression, c) without prejudice to Union or national or Union legislation on copyright, document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law."

[8] See Article 52(3), "Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated."

[9] See https://publications.parliament.uk/pa/ld5803/ldselect/ldcomm/125/12505.htm#_idTextAnchor019

5. *Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.*

We urge the Copyright Office to confirm that it shares our view that, under U.S. law, the use of copyrighted works to train commercial Generative AI Models and AI Systems requires authorization from the relevant rightsholders. If courts were to interpret existing law as permitting the widespread use of copyrighted content to train Generative AI models without rightsholder consent, we would support new legislation to ensure that consent is required.

We also support regulation that mandates transparency requirements and obligates the developers of AI Models and AI Systems to:

- Follow transparency standards for training data: AI Models should only be trained on authorized data and AI developers should be obligated to retain auditable records that document training data used for Machine Learning. In the context of Generative AI, it is especially important that such records include information related to the use of protected intellectual property as training data and the sources from which such works were obtained. Knowing the sources from which the works have been copied is important as often unlicensed copyrighted material used as training data is scraped from known piracy sites or in violation of the terms of use of an otherwise legitimate source of the content.

- Label Generative Output: Generative AI Systems must clearly, specifically, and consistently identify their outputs and interactions as including AI Generated content and the identity of the applicable AI Model.

Additionally, we support regulation that obligates AI developers and providers of AI Services to share in the responsibility to address misinformation, bias and misuse embodied in or created by such AI Services or the underlying AI Models.

Training

6. *What kinds of copyright-protected training materials are used to train AI models, and how are those materials collected and curated?*

A wide variety of copyright-protected works, including images, videos, fine art, text and music, have been used to train AI Models, although some AI Models focus on a particular type of content. In the context of Machine Learning that requires visual data, high quality visual content that is digitized, organized and accompanied by reliable and comprehensive caption information and metadata is particularly valuable.

Getty Images' growing library of over 538 million visual assets (images and video) is well suited for AI training purposes and large sections of our portfolio have been used in data training sets for Generative AI Models, in some cases with our authorization and, in at least one prominent example described in detail in our Amended Complaint, without our consent and in flagrant violation of our

intellectual property rights and the terms of service of the websites from which our works were scraped.

Getty Images and other similar organizations that license their content as training sets for AI Models typically curate and tailor the dataset for the licensee's desired use. For instance, where a licensee is developing a text-to-image model for commercial use, Getty Images curates a dataset that includes content that has been released for commercial use in respect of rights of publicity, privacy, trademark and other intellectual property rights. See response to Question 6.2 for more detail on the market for licensed datasets.

In many instances, however, as explained below, some developers of AI Models have included copyrighted works in training sets without obtaining permission from the relevant rightsholders. In some of those cases, the copyrighted works were scraped from internet websites in violation of express terms of use; in others, copies of the copyrighted works were obtained from known piracy sites; and in others, the model developer may have had authorized access to the copyrighted works for a different purpose but used the works as training data notwithstanding express contractual limitations on such uses.

6.1 How or where do developers of AI models acquire the materials or datasets that their models are trained on? To what extent is training material first collected by third-party entities (such as academic researchers or private companies)?

Further to the information provided in response to Question 6, datasets for Generative AI Models may be acquired from a variety of sources, including but not limited to licensors who have the right to grant sublicenses for AI training purposes (please see the response to Question 6.2 for more detail on licensed datasets); compiled by the AI developer itself, often by scraping the internet for content without consent of the rightsholders; or from third-party research entities affiliated with institutions of higher learning or non-profits.[10]

A number of datasets have been created by research entities by scraping the internet and including content without the consent of the rightsholder, ostensibly for non-commercial use but often with funding and other support from for-profit entities that intend to, and do, then make commercial uses of the datasets. For-profit technology companies that make subsequent uses of those datasets in some cases have attempted to justify their unauthorized exploitation of copyrighted works under the pretense that it is "fair use" to use the copyrighted works as training data because the copying was facilitated by a non-profit entity. For-profit organizations cannot rely on non-commercial organizations to "clean" datasets of obligations

[10] For example, Meta, in training its Make-A-Video text to video AI generation tool, in addition to using 10.7m clips scraped from Shutterstock, also used 10m video clips from the Microsoft Research Asia data set from which all of the clips came from YouTube videos. See also for example, https://openaccess.thecvf.com/content/CVPR2022/supplemental/Xue_Advancing_High-Resolution_Video-Language_CVPR_2022_supplemental.pdf The image-text datasets used to train Stability AI's Stable Diffusion and Google's Imagen were reportedly first compiled by LAION, a German entity that has been supported by Stability AI.

to acquire consent. This phenomenon has been discussed by the media and is often referred to as "data laundering"[11].

6.2 To what extent are copyrighted works licensed from copyright owners for use as training materials? To your knowledge, what licensing models are currently being offered and used?

There are available, established paths for licensing large sets of high-quality visual works and associated metadata for use in connection with the training and development of AI Models, and responsible AI developers are licensing those datasets from providers such as Getty Images, Envato, Alamy and Shutterstock. Fully licensed, commercially safe, and non-infringing products, such as Generative AI by Getty Images and Firefly by Adobe, capable of generating high-quality images from text prompts are already available in the marketplace.

The ecosystem for licensing visual works as parts of training datasets promotes a scaled source of demand for copyright protected work that enables human creators to be compensated when Generative AI Systems are commercialized or compete with those creators' own efforts. That demand helps incentivize creators to produce a continuous source of up-to-date and organized content and metadata that satisfies the needs of developers of Generative AI Models and AI Systems and the needs of the marketplace for visual works more generally.

6.3 To what extent is noncopyrighted material (such as public domain works) used for AI training? Alternatively, to what extent is training material created or commissioned by developers of AI models?

There is a vast array of public domain materials available to AI developers who wish to train AI Models without incurring licensing fees or infringing third-party rights, but AI developers appear to prefer using copyrighted materials likely because of the generally superior depth, breadth and quality that they provide over content that is in the public domain.

Additionally, and importantly for the creation of responsible AI Models, public domain materials do not come with any assurances that they are suitable for commercial use. The subjects of public domain images and videos often include individuals who have not signed images/likeness releases, properties for which the owners have not signed property releases and representations of brands, trademarks and products. Similarly, in our experience, databases of public domain materials are often vetted incorrectly or not vetted at all and may wrongfully include copyrighted works, leaving it to the copyright owner to self-police.

6.4 Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.

[11] See https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/

Retention practices among AI developers may vary, but the extent to which training materials are retained after the model is trained does not bear on whether the extensive copying of those materials during the training process is infringing.

7. *To the extent that it informs your views, please briefly describe your personal knowledge of the process by which AI models are trained. The Office is particularly interested in:*

7.1. How are training materials used and/or reproduced when training an AI model? Please include your understanding of the nature and duration of any reproduction of works that occur during the training process, as well as your views on the extent to which these activities implicate the exclusive rights of copyright owners.

There are many techniques used for training AI Models on copyright protected works. Diffusion models are widely used in the context of Generative AI and can be used for image generation.[12] In our understanding of at least one example, to train a diffusion model, an AI developer first loads a copy of training data files (images) into computer memory. Next, they encode the images, which involves creating smaller versions of the images that take up less memory. In addition, the AI developer also encodes text that describes each image. The AI developer may retain and store copies of the encoded images and text as an element of training the AI Model. Third, the AI Developer adds visual "noise" to the encoded images, i.e., it further alters the images so that it is incrementally harder to discern what is visually represented because the images have been intentionally degraded in visual quality in order to "train" the AI Model to remove the "noise." By intentionally adding visual noise to the existing images with associated text, the AI developer teaches the AI Model to generate output images to be consistent with a particular text description. Fourth, the AI Model decodes the altered image and teaches itself to remove the noise by comparing the decoded image to the original image and text descriptions that have been copied and stored. By learning to decode noise, the AI Model in this example learns to deliver images similar to—and, in some cases, substantially similar to—the original without noise.[13]

7.2. How are inferences gained from the training process stored or represented within an AI model?

Because of the inherent complexity of AI Models, the details of how inferences are gained and stored are not fully understood. Some AI developers choose to address the unknown by making claims that AI Models learn in the same way that humans do. This is a misleading assertion that ignores the fact that, unlike with human learning, computer learning requires that a digital copy of expressive work be copied (and typically copied numerous times) in order to utilize it for training. Machine learning is intrinsically different from human learning and the two should not be conflated. Unlike in human learning, it is not possible to assert that there is a separation between an inference and copyrightable expression once a protected work has

[12] See https://en.wikipedia.org/wiki/Diffusion_model

[13] See Amended Complaint in Getty Images (US), Inc v Stability AI, LTD and Stability AI, Inc for an example of an AI Developer who trained a diffusion Generative AI Model: https://docs.justia.com/cases/federal/district-courts/delaware/dedce/1:2023cv00135/81407/13

been used to train a model. While the output of Generative AI systems may not always be identical to any particular individual work used for training, studies have shown that diffusion models used to generate synthetic imagery are capable of memorizing and reproducing copies of images that were included in the training set.[14]

7.3. Is it possible for an AI model to "unlearn" inferences it gained from training on a particular piece of training material? If so, is it economically feasible? In addition to retraining a model, are there other ways to "unlearn" inferences from training?

The concept of "unlearning" or "forgetting" specific inferences in an AI Model has been a subject of ongoing research. While some studies have demonstrated the feasibility of this process, it is typically limited to erasing a single or a small number of concepts from the applicable AI Model's knowledge base.

One of the inherent challenges in unlearning is that concepts within AI Models are intricately intertwined. This means that attempting to remove or alter one concept can inadvertently affect other related or even seemingly unrelated concepts. For instance, if the AI Model were to unlearn information about specific individuals like "Barack Obama" or "Donald Trump," this process could potentially distort the broader concept of "U.S. president," as these individual instances may be deeply interconnected within the AI Model's parameter space. Similarly, unlearning information gleaned from an unauthorized copyrighted work (such as any attempt that could be made in an after the fact opt out process) is unlikely to be successful given the interdependence on other concepts within the AI Model.

Furthermore, considering the vast number of concepts an AI Model learns during training, unlearning at a large scale becomes effectively impossible. Attempting to do so would likely result in significant degradation of the AI Model's performance and utility, as countless interrelated concepts would be impacted. In many cases, it could be more practical to retrain the AI Model completely.

From an economic standpoint, the process of unlearning is not feasible when considering the scale of retraining or modification required to achieve it without harming the AI Model's integrity.

7.4. Absent access to the underlying dataset, is it possible to identify whether an AI model was trained on a particular piece of training material?

No response.

8. *Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.*

[14]See Extracting Training Data from Diffusion Models, Nicholas Carlini, Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr, Borja Balle, Daphne Ippolito, Eric Wallace at https://arxiv.org/abs/2301.13188

The fair use doctrine embodied in Section 107 of the Copyright Act sets forth four non-exclusive factors that courts must consider in assessing whether an otherwise infringing act should be excused as a fair use of copyrighted material.[15] As the Copyright Office is of course well aware, these factors include: (i) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (ii) the nature of the copyrighted work; (iii) the amount and substantiality of the portion used; and (iv) the effect of the use upon the potential market for or value of the copyrighted work. While the list of factors is non-exclusive, courts must consider each of the statutory factors and weigh them in light of the purposes of the Copyright Act.[16] To prevail on a fair use defense, "a defendant must convince the court that allowing his or her unpaid use of copyrighted material would be equitable and consonant with the purposes of copyright."[17]

The application of the fair use doctrine "always depends on consideration of the precise facts at hand."[18] We address several of the statutory factors in more detail in response to specific subsections below.

8.1. In light of the Supreme Court's recent decisions in Google v. Oracle America and Andy Warhol Foundation v. Goldsmith, how should the "purpose and character" of the use of copyrighted works to train an AI model be evaluated? What is the relevant use to be analyzed? Do different stages of training, such as pre-training and fine-tuning, raise different considerations under the first fair use factor?

The first statutory factor of the fair use analysis requires courts to consider the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes. As the Supreme Court recently explained in *Warhol*, the first statutory factor "considers the reasons for, and nature of, the copier's use of an original work."[19] The central question is "whether the new work merely supersedes the objects of the original creation . . . (supplanting the original), or instead adds something new, with a further purpose or different character."[20] Describing "the problem of substitution" as "copyright's bête noire," the Court observed that "[t]he use of an original work to achieve a purpose that is the same as, or highly similar to, that of the original work is more likely to substitute for or supplant the work.[21] Accordingly, "[i]f an original work and a secondary use share the same or highly similar purposes, and the secondary use is of a commercial nature, the first factor is likely to weigh against fair use, absent some other justification for copying.[22] Whether an allegedly infringing use has a further

---

[15] 17 U.S.C. § 107

[16] See Andy Warhol Found. For the Visual Arts, Inc. v. Goldsmith, 143 S. Ct. 1258, 1287 (2023) ("Warhol").

[17] Cambridge Univ. Press. v. Patton, 769 F.3d 1232, 1238 (11th Cir. 2014).

[18] Am. Geophys. Union v. Texaco Inc., 60 F.3d 913, 916 (2d Cir. 1994); see also Warhol, 143 S. Ct. at 1274 ("[F]air use is "a flexible concept, and its application may well vary depending on context.") (quotation omitted); Fox News Network, LLC v. TVeyes, Inc., 883 F.3d 169, 176 (2d Cir. 2018) (noting "case-by-case" nature of the fair use analysis)

[19] Warhol, 143 S. Ct. at 1274.

[20] Id. (quotations omitted)

[21] Id. (quotations omitted)

[22] Id. at 1277; see also NXIVM Corp. v. Ross Inst., 364 F.3d 471, 481 (2d Cir. 2004) (observing that when a defendant "offer[s] a market substitute for the original," its use is not fair).

purpose or different character is a matter of degree, and the degree of difference must be weighed against other considerations, like commercialism.[23]

While it is conceivable that copyrighted material could be used in connection with training an AI Model for a completely different purpose than the purpose the original work serves, developers in many cases have used copyrighted materials without authorization to train AI Models and AI Systems designed to serve the same purpose as the original works on which they have been trained. There is, for example, nothing fair about mass copying of copyrighted recordings of songs by a popular performer to train a model to generate additional recordings that sound like that performer's original works and are distributed to entertain listeners no differently than the original works. And there is nothing fair about using copies of millions of copyrighted photographs and the associated metadata scraped from the website of a leading distributor and licensor of photographs to train a model to generate the same kinds of images and then offering those models to provide customers with alternative source to obtain those same kinds of images.[24]

Whether the use of copyrighted materials in a training dataset is for commercial or nonprofit educational purposes is a relevant consideration to the determination of fair use[25], but the answer to that question is not dispositive. While many commercial uses may, on balance, be fair, the Supreme Court has long recognized that commercial use of copyrighted works weighs against fair use.[26] Many nonprofit educational uses of copyrighted materials, however, do not qualify as fair uses either.[27] At bottom, "[t]he crux of the profit/nonprofit distinction is ... whether the user stands to profit from exploitation of the copyrighted material without paying the customary price."[28] The fair use doctrine "distinguishes between a true scholar and a chisler who infringes a work for personal gain."[29]

The application of these principles to the use of copyrighted materials to train generative AI Models is relatively straightforward. When done at scale for commercial purposes, such copying is presumptively unfair. That is true whether the training itself is performed directly by employees of

---

[23] Warhol, 143 S. Ct 1276.

[24] The unfairness of these acts is compounded when the copies are obtained in violation of the terms of use of the websites from which they are scraped or when they are obtained from a source that is itself infringing, such as an online repository of pirated copies of books or music.

[25] see 17 U.S.C. § 107(1),

[26] See Harper & Row, 471 U.S. at 562 ("'[E]very commercial use of copyrighted material is presumptively an unfair exploitation of the monopoly privilege that belongs to the owner of the copyright.'") (quoting Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417, 451 (1984)); see also, e.g., Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191, 198 (3d Cir. 2003) ("If a new work is used commercially rather than for a nonprofit purpose, its use will less likely qualify as fair."); Am Geophys. Union, 60 F.3d at 022 ("[C]ourts will not sustain a claimed defense of fair use when … the copier directly and exclusively acquires conspicuous financial rewards from its use of the copyrighted material.").

[27] See, e.g., Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584 (1994) ("[T]he mere fact that a use is educational and not for profit does not insulate it from a finding of infringement."); Soc'y of Holy Transfiguration Monastery, Inc. v. Gregory, 689 F.3d 29 (1st Cir. 2012) (holding that archbishop's fair use defense for copying translation of religious text failed in full, notwithstanding nonprofit education nature of use); Princeton Univ. Press v. Michigan Documents Servs., Inc. 99 F.3d 1381, 1390-1392 (discussing Agreement on Guidelines for Classroom Copying in Not-for-Profit Educational Institutions With Respect to Books and Periodicals in legislative history of Section 107).

[28] Harper & Row, 471 U.S. at 562.

[29] Id. at 563.

a for-profit entity or by researchers at nonprofit institutions who then make their AI Models and the output from those AI Models available for commercial uses (often with funding or other support from commercial entities who will profit from those models). When done for noncommercial purposes, such copying will still not be fair if the uses interfere with established licensing markets, such as the licensing of high-quality copyrighted works for use in training sets, or supersede the use of the original works, as is the case, for example, with text-to-image models that offer substitutes for stock photography and other commercial markets for images.

The same analysis that is applied to training the initial AI Model should be considered with respect to other stages of training, such as pre-training and fine tuning.

8.2. How should the analysis apply to entities that collect and distribute copyrighted material for training but may not themselves engage in the training?

The fair use analysis applies no differently to entities that collect and distribute copyright material for training than it applies to entities that conduct the training. In each case, the entity has violated at least one of the exclusive rights reserved to copyright holders under Section 106 of the Copyright Act and is liable for infringement absent an applicable defense.[30] In addition, entities and individuals that contribute to or induce the infringement or have the right and ability to supervise the infringing activity and direct financial interest in it may be held secondarily liable for infringing acts, even if they do not themselves engage in acts that violate one of the exclusive rights under Section 106.

8.3. The use of copyrighted materials in a training dataset or to train generative AI models may be done for noncommercial or research purposes. How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

As noted in response to Question 8.1, the distinction between commercial and noncommercial purposes is relevant to the fair use analysis but not dispositive. And the distinction becomes unimportant when activities conducted for ostensibly noncommercial purposes are subsequently adapted for commercial purposes or when funding for noncommercial or research uses is supplied by for-profit entities that seek commercial benefits from those activities. Please see our response to Question 6.1 regarding data laundering.

Even if all of the entities involved in the development and distribution of a Generative AI Model are noncommercial, the use of copyrighted works may still not be fair if they are interfering with established or likely licensing markets and/or offering competing products without having invested in the creation of the copyrighted works on which the models are trained.

8.4. What quantity of training materials do developers of generative AI models use for training? Does the volume of material used to train an AI model affect the fair use analysis? If so, how?

[30] See 17 U.S.C. 106 (providing copyright owners with exclusive rights to reproduce their copyrighted works and to distribute those works, among other rights).

The third statutory fair-use factor requires courts to consider the amount and substantiality of the portion used. While copying the entire work does not preclude a finding of fair use when all four factors are balanced against each other, "copying an entire work militates against a finding of fair use."[31]

The volume of material taken from a particular copyright owner may impact the effect on the market for that owner's works and thus be relevant to the fair use analysis, but the total volume of materials used to train an AI model is unlikely to affect the fair use analysis. As the Supreme Court has recognized, "'no plagiarist can excuse the wrong by showing how much of his work he did not pirate.'"[32]

The quantity of training data needed varies depending on the type of AI Model and its desired function. Some AI developers have argued that the quantity of copyrighted works needed to train their Generative AI Models made it impossible to consider licenses of such content.[33] This is simply untrue. In the context of visual machine learning, the quantity of visual work used for training may be less important than its quality. Getty Images and Adobe have both released AI image generation tools supported by AI Systems and AI Models that were trained solely on permissioned content (licensed content and, in the case of Adobe, some public domain content).

8.5. Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured? Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?

The fourth statutory factor requires courts to consider "the effect of the use upon the potential market for or value of the copyrighted work."[34] While competition between outputs of an AI Model and a particular copyrighted work used to train the AI Model, the body of works of the same author, or the market for that general class of works may all be relevant considerations under the fourth statutory factor, the analysis is not properly limited to assessing competition from the outputs of a particular AI Model. Rather, the inquiry is considerably broader in at least two important ways. First, the fourth statutory factor requires courts "to consider not only the extent of the harm cause by the particular actions of the alleged infringer, but also whether unrestricted and widespread conduct of the sort engaged in by the defendant would result in a substantially adverse impact on the potential market for the original."[35] Second, the inquiry must also consider

---

[31] VHT, Inc. v. Zillow Group, Inc., 918 F.3d 723, 744 (9th Cir. 2019) (quotation omitted); see also 4-13 Nimmer on Copyright § 13.05 ("In general, it does not constitute a fair use if the entire work is reproduced.").
[32] Harper & Row, 471 U.S. at 566 (quoting Sheldon v. Metro-Goldwyn Pictures Corp., 81 F.2d 49, 56 (2d Cir. 1936)).
[33] https://www.digitalcameraworld.com/news/midjourney-founder-basically-admits-to-copyright-breaching-and-artists-are-angry
[34] 17 U.S.C. § 107(4).
[35] Campbell, 510 U.S. at 590; see also, e.g., Andy Warhol Found. for Visual Arts, Inc. v. Goldsmith, 11 F.4th 26, 49-50 (2d Cir. 2021), aff'd, 143 S. Ct. 1258 (2023).

whether, in using copyrighted works to train an AI Model, the alleged infringer has exercised one of the rights reserved to the copyright holder "without paying the customary price."[36]

As described in our response to Question 6 above, there is an established path available for licensing copyrighted works for use in training generative AI Models. Accordingly, the loss of license revenue from that market (whether existing revenues, potential revenues, or both) is cognizable market harm under the fourth factor, even in the absence of competing uses of model outputs. As the Second Circuit recognized in American Geophysical Union v. Texaco Inc., "[i]t is indisputable that, as a general matter, a copyright holder is entitled to demand a royalty for licensing others to use its copyrighted work, see 17 U.S.C. § 106 (copyright owner has exclusive right 'to authorize' certain uses), and that the impact on potential licensing revenues is a proper subject for consideration in assessing the fourth factor."[37] In rejecting Texaco's fair use defense, the Second Circuit appropriately recognized the market harm resulting from the unauthorized copying of scientific journal articles by researchers at Texaco, even though the copies made were not sold or distributed in competition with the original works, the products created and sold by Texaco did not compete with the journals copied, and more than half of the articles copied without permission at issue were never used for research at all.[38]

Similarly, in Fox News Network, LLC v. Tveyes, Inc., the Second Circuit held that Tveyes' use of copyrighted Fox newscasts in connection with a searchable video clip service was infringing, even though the service posed little risk of being a competing substitute for Fox's offerings.[39] Rather, "the success of Tveyes' business model demonstrate[d] that deep-pocketed consumers [were] willing to pay well" for the service, and Tveyes' unauthorized use was depriving Fox of the opportunity to license its content to Tveyes or a similar service.[40] So too with the use of copyrighted works in the training sets of AI Models. The unauthorized use of copyrighted works in training sets for AI models both provides substantial value to those who develop and exploit the models and deprives copyright holders of licensing revenues. Depriving copyright holders of the opportunity to license their works for use in training sets is by itself sufficient market harm to cause the fourth statutory factor to weigh against fair use.

9. *Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?*

Copyright is an opt-in regime and those who seek to exploit one of the rights exclusively reserved to the copyright holder under Section 106 need to obtain affirmative permission to do so. Establishing an opt-out regime would require a change in U.S. copyright law. For the reasons set

---

[36]N .Harper & Row, 471 U.S. at 562; see also Associated Press v. Meltwater U.S. Holdings, Inc., 931 F. Supp. 2d 537, 553 (S.D.N.Y. 2013) ("Meltwater") ("Where there is a fully functioning market for the infringer's use of the copyrighted material, it will be difficult for the infringing party to show that it made fair use without paying a license fee.")

[37] 60 F. 3d 913, 929 (2d Cir. 1994); see also Princeton Univ. Press, 99 F.3d at 1387-88.

[38] Id. at 915-16.

[39] 883 F.3d at 179-180.

[40] Id. At 180; see also Meltwater, 931 F. Supp. 2d at 553 (rejecting fair use defense because, inter alia, "permitting Meltwater to avoid paying license fees gives it an unwarranted advantage over its competitors who do pay licensing fees").

out below in response to question 9.2, including that an opt out after the fact is impractical as the value has already been conveyed and retained and we do not support such a change in the law.

9.1. Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses?

Pursuant to existing U.S. copyright law, consent of the rightsholder is required for all use of copyrighted materials as AI training data unless a defense (such as fair use) applies and this application of the law should be upheld. See response to Question 8 for further discussion of fair use.

9.2. If an "opt out" approach were adopted, how would that process work for a copyright owner who objected to the use of their works for training? Are there technical tools that might facilitate this process, such as a technical flag or metadata indicating that an automated service should not collect and store a work for AI training uses?

An "opt-out" approach to AI training is not a suitable approach to consent for a number of reasons, including that:

- it does not protect or respect rightsholders,
- it does would not address learning/ training by AI Models prior to opt-out;
- it places an unfair burden on creators to know where and how their works have been used absent disclosure requirements and to navigate inconsistent processes across a large and growing volume of AI Services; and
- it does not work for large portfolios of work (Getty Images, for example, owns or represents more than 538 million copyrighted works).

Some AI developers have taken the position that copyright owners should use robot.txt to pre-identify content that should be opted out of training. Robot.txt is not a viable means of opt-out. Robot.txt is applied at a webserver (i.e., website) level versus the content level. Consistent with many content owners, display of Getty Images' content is not limited to our websites. It is distributed across the millions of websites and social media where Getty Images and other creators cannot mandate the deployment of robot.txt.

Rather than demand a new notice is attached to copyrighted works opting out of training, current IPTC metadata regularly attached to visual data identifies key attributes of visual work including "Copyright Notice" which identifies the copyright owner and "Credit" which identifies the distributor. Technical tools are readily available that enable humans and machines to surface such metadata and there should be a presumption that authorization is needed, especially, but not only, if a copyright notice exists. One limitation of the current methods of attaching IPTC metadata to copyright protected content is that such metadata is easily stripped however the impact of stripping such metadata would be limited by enacting transparency obligations that promote the retention of such data. See response to Questions 15 and 15.1 for more detail on obligations of transparency.

9.3. What legal, technical, or practical obstacles are there to establishing or using such a process? Given the volume of works used in training, is it feasible to get consent in advance from copyright owners?

It is certainly feasible to get consent for AI training in advance. Responsible developers of Generative AI Models and AI Systems are licensing training datasets from Getty Images, Envato, Alamy, Shutterstock and other similar organizations. As set out in our response to Question 6 above, there is an established path for licensing visual data for use in training, and there are already fully licensed, non-infringing products in the marketplace.

In addition to our response to Questions 7.3 and 9.2, a major practical obstacle to an "opt-out" process is that, even if a copyright owner does opt out after they become aware of a Generative AI Model that has been trained on their work, the opt-out cannot erase the impact on the training that has already been done, where both the impacted AI Model and its outputs may have been commercialized. Unlearning is not effective and retraining a Generative AI Model from scratch each time an opt-out is received is neither practical nor the practice of Generative AI Model developers who have established opt-out processes.

In addition to our response to Question 9.2, applying a "do not train" message in a site's robots.txt is not a practical solution for rightsholders because it may block crawling for more than just training purposes and could prevent indexing of the websites of rightsholders by search engines that the rightsholders find desirable, limiting their ability to be found on the web and causing damage to established business models. Moreover, taking such steps would likely be ineffective in any event if the copyrighted work in question is published online on different websites, as is the case for images and video clips licensed by Getty Images to its customers.

9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?

As described above, "opt-out" is not required by copyright holders to reserve their rights under U.S. copyright law. While we disagree with any amendment to U.S. copyright law that would shift from an opt-in system of consent to an opt-out system, if the law were changed, the remedies available under current law for copyright infringement should also be available for failure to honor an opt-out.

9.*5. In cases where the human creator does not own the copyright—for example, because they have assigned it or because the work was made for hire— should they have a right to object to an AI model being trained on their work? If so, how would such a system work?*

We do not believe that copyright law should be expanded such that creators who assigned their copyrights or who provided their services in connection with works made for hire would maintain a non-assignable right of consent for AI training purposes.

10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?

Licenses to use copyrighted works to train Generative AI Models should be obtained in the way most copyright licenses are obtained: through bilateral negotiations between the licensee and the respective owners of the content they wish to license. For some types of content, there may be aggregators who are authorized to license pools of content on behalf of multiple owners. Licenses to scaled quantities of content and metadata required to train Generative AI Models are already readily available in the marketplace. The claim by some developers that there is no way to get consent from copyright holders given the quantity of materials needed to train AI Models[41] is simply untrue. There is an established market for training data, and there is a growing body of high-quality Generative AI Models that have been trained on content licensed for that purpose.

10.1. Is direct voluntary licensing feasible in some or all creative sectors?

Yes, as described above in our response to Questions 6 and 10. We are not aware of any creative sector for which direct licensing would not be feasible.

10.2. Is a voluntary collective licensing scheme a feasible or desirable approach? Are there existing collective management organizations that are well-suited to provide those licenses, and are there legal or other impediments that would prevent those organizations from performing this role? Should Congress consider statutory or other changes, such as an antitrust exception, to facilitate negotiation of collective licenses?

While one or more voluntary collective licensing schemes may be feasible, not all visual content is the same. Getty Images prefers to license its copyrights directly. As described in our response to Question 10, in the context of datasets that contain visual work, a market does exist for quality content that facilitates direct licensing. That said, even when licensing directly, an AI developer may wish to source training data from a number of different licensors and there could be instances where it is in the wider public interest for dataset owners to be able to collaborate with regard to licensing structures and terms, to encourage sustainable licensing arrangements over the medium to long term and to enable the inclusion of a more diverse range of content to be used as training data. In this regard, it would be helpful for the appropriate anti-trust authorities to issue guidance regarding the level of collaboration amongst copyright holders who wish to license collectively in this context that is permitted under existing anti-trust laws.

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?

See answer to 10.4 below.

10.4. Is an extended collective licensing scheme a feasible or desirable approach?

[41] See, https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=571764c32d2b

To answer both questions 10.3 and 10.4, compulsory or extended collective licensing schemes are not desirable when a marketplace for direct licensing already exists, which is the case with the licensing visual works and metadata to use in connection with the training and development of AI Models. Extended collective licensing may be a solution to lower transactions costs when dealing with independent artists or licensors of small numbers of copyrights, however in this context it is not necessary to promote AI innovation. In the event that policy makers see it fit to enact an extended collective licensing scheme, there needs to be an easy and efficient mechanism that rightsholders can use to opt out of the scheme if they prefer not to participate.

10.5. Should licensing regimes vary based on the type of work at issue?

Perhaps. As described above in our response to Question 10, the dataset markets for different types of work are unique and have developed independently.

11. *What legal, technical or practical issues might there be with respect to obtaining appropriate licenses for training? Who, if anyone, should be responsible for securing them (for example when the curator of a training dataset, the developer who trains an AI model, and the company employing that model in an AI system are different entities and may have different commercial or noncommercial roles)?*

There are no practical or technical issues that justify the unauthorized use of copyrighted works as training data or the commercial use of AI Models or AI Services trained on copyrighted works without consent of the relevant rightsholders.

As with obligations for consent from copyright owners for non-AI related uses of copyrighted works, any party that is using copyrighted materials for training purposes or providing an AI Service based on an AI Model developed with copyrighted materials must obtain consent from the applicable rightsholders.

While it may be more convenient and more economical to the developers of AI Models and AI Services to use unauthorized copyrighted works for free rather than obtain licenses and pay licensing fees, there is no credible argument that licensing costs will inhibit innovation. The multi-billion-dollar scale of investment that leading technology companies have made in developing AI Systems and AI Models[42] accommodates the cost of obtaining licenses and there is no reason to believe that respect for copyright laws in the context will inhibit innovation. Innovation and intellectual property laws coexist in the technology sector more broadly and will do so here. Responsible Generative AI Model developers such as Getty Images, NVIDIA, Bria and Adobe have

[42] See https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html See also for example, Amazon's recent announcement of $3B commitment to leading Generative AI developer Anthropic. https://www.aboutamazon.com/news/company-news/amazon-aws-anthropic-ai, Microsoft's $10bn investment in Open AI earlier this year: https://www.bloomberg.com/news/articles/2023-01-23/microsoft-makes-multibillion-dollar-investment-in-openai, and Meta' announcement to invest a colossal $33bn in AI in2023: https://www.thestack.technology/meta-ai-investment

proven that it is possible to compensate creators for the use of copyright protected works and develop effective, commercially safe Generative AI Models.

12. *Is it possible or feasible to identify the degree to which a particular work contributes to a particular output from a generative AI system? Please explain.*

No response.

13. *What would be the economic impacts of a licensing requirement on the development and adoption of generative AI systems?*

We believe that the more important question to ask is: what is the cost of not requiring licensing? As further described in our response to Question 1, allowing development and deployment of commercial AI Models and AI Services trained on unlicensed content, will cause a negative economic impact for creatives, the free press, and society at large. Developers of AI Models and AI Systems should not be permitted to usurp existing and potential markets for copyrighted works.

Rightsholders benefit from the positive economic impact and the financial incentive of licensing their works for training AI Models and AI Systems, allowing for sustainability of the creative economy. It is unfair, unnecessary, and unjustified to transfer this benefit to the developers of AI Models who use copyright protected materials without authorization. As part of factor 4 of the fair use analysis, explicit consideration of "the effect of the use upon the potential market for or value of the copyrighted work" is required. Where the output of an AI Model competes with the copyrighted content used to train the model, there is the potential for a significant impact on the market for the copyrighted work. The licensing requirement ensures respect for the value of copyrighted work and incentivizes human creativity.

As set out in response to Question 11, the costs associated with obtaining authorization to use copyrighted works should be viewed as a reasonable cost of innovation, as shown by Getty Images, Adobe and Bria, among others. Innovation and respect of intellectual property laws coexist comfortably in other technology sectors, and there is no reason that they cannot do it in this context. The fair use doctrine does not allow defendants to avoid the customary price of a license just because they cannot or prefer not to pay for one.

14. *Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models*

The role that open-source AI Models play in the development of AI Systems needs to be considered when thinking about potential copyright liability. In particular, in the context of Generative AI, open-source foundation models that have been trained on massive amounts of copyright protected works can be freely used to develop sophisticated Generative AI Systems. In such cases, it is essential that the developers of such AI Systems are not able to rely on the fact that the underlying AI Model is open-source in order to avoid liability for the downstream consequences of their copyright infringement that results from others using and adapting their models if their models have been trained on copyrighted works without securing license authority from the

relevant rightsholders. Any laws or policies that would exempt open-source AI Models from the same legal or regulatory obligations imposed on non-open-source AI Models should be avoided. [43]

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?

Developers of AI Models and creators of training datasets should be required to collect, retain and disclose auditable records regarding the sources of materials used to train AI Models. Because AI developers are not currently required to disclose the dataset(s) upon which they have trained their AI Models and AI Systems most refuse to do so, even when requested. For example, Midjourney CEO, David Holz, has publicly stated, "Our training data is pretty much from the same place as everybody else's – which is pretty much the internet. Pretty much every big AI model just pulls off all the data it can, all the text it can, all the images it can."[44] Yet, notwithstanding this claim, in response to a claim for copyright infringement for the use of copyrighted works as training data without consent of the rightsholders, Midjourney has responded in its Motion to Dismiss in *Andersen et al. v. Stability AI, Ltd.,* that the Plaintiffs carry the burden of identifying the specific work that were used without consent.[45] The "heads I win, tails you lose" approach stacks the deck in favor of infringers.

15.1. What level of specificity should be required?

All AI developers should be required to keep confidential detailed auditable records and at a minimum, such auditable records should include:

- Disclosure of all public and private datasets that are likely to contain copyright protected work and have been used to train Generative AI Models. Disclosure should include records of the name of the datasets used and a short description of each. Where a dataset includes unlicensed scraped content, the URL from where each item of content was scraped should also be disclosed.

- Disclosure of when copyrighted protected works are being used for AI/ML training purposes.[46]

---

[43] The concerns surrounding open-source AI models were also raised during the Senate Judiciary Committee, Subcommittee on Privacy, Technology, and the Law, hearing on Oversight of A.I.: Principles for Regulation, during which witnesses testified to the dangers of opening up the operation of generative AI technologies to bad actors. See https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-principles-for-regulation.

[44] https://www.theverge.com/2022/8/2/23287173/ai-image-generation-art-midjourney-multiverse-interview-david-holz

[45] See for example, Motion to Dismiss in *Andersen v Stability AI* Case No. 3:24-cv-00201-WHO

[46] See for example, Article (28b(4)(c) of European Parliament Proposed text for the AI Act
"Providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("generative AI") and providers who specialise a foundation model into a generative AI system, shall in addition
a) comply with the transparency obligations outlined in Article 52 (1),

- Information sufficient to identify the entitlement to use the works in the dataset (e.g., ownership, license, public domain)

15.2. To whom should disclosures be made?

The records described in our response to Question 15.1 should be made available in an open and searchable fashion.

15.3. What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?

The ordinary rules of direct and secondary liability for copyright infringement under the Copyright Act law should apply equally to developers of AI Systems that incorporate AI Models from third parties as it does to those that develop the AI Models themselves.

15.4. What would be the cost or other impact of such a recordkeeping system for developers of AI models or systems, creators, consumers, or other relevant parties?

The costs of implementing a recordkeeping system of the nature set out in response to Questions 15 and 15.1 should be borne by developers of AI Models. As set out in response to Question 11 regarding the cost of licensing, the costs associated with recordkeeping will not be an impediment to innovation.

16. *What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?*

The obligation on the part of AI developers is to get a license from copyright owners before using their works to train AI Models, not to give mere notification after using without authorization. If no license is needed (public domain, or Creative Commons), no notice is necessary however see our response to answer to Question 15 above in respect of recordkeeping.

17. *Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?*

No response.

Generative AI Outputs

Copyrightability (Q. 18 – 21)

b) train, and where applicable, design and develop the foundation model in such a way as to ensure adequate safeguards against the generation of content in breach of Union law in line with the generally acknowledged state of the art, and without prejudice to fundamental rights, including the freedom of expression,
c) without prejudice to Union or national or Union legislation on copyright, document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law."

Our views in respect of the responses to Questions 18 and 21 are consistent with the views expressed in the responses submitted by the Copyright Alliance. In addition, in respect of Question 20, we do not believe that legal protection specifically for AI generated material is necessary to encourage development of Generative AI Models and AI Systems. Such AI Models and AI Systems are already being developed at pace absent any specific legal protection of their outputs.

Infringement (Q. 22-27)

For more information, please see the Amended Complaint, Getty Images (US), Inc. v. Stability AI, Inc., attached as ***Appendix A*** hereto. We reserve further public comment on the issues addressed in Questions 22-27 at this time.

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?

The output of Generative AI Systems should be required by law to be identified as such in a clear, consistent and persistent manner. Current protocols that may be easily stripped from an image such as IPTC, c2Pa, and digital watermarks, are not adequate solutions as further described in our response to Question 28.3 below. Embedded metadata is currently nowhere close to indelible and is frequently removed by users. The burden of identification should remain primarily with the providers of AI Models and AI Systems.

Generative AI identification solutions must be:

1. Close to Indelible

Publication of imagery and video regularly involves workflows that manipulate both the pixels and metadata of media files, resulting in metadata stripping. Any solution for identifying Generative content must reliably classify content after it has gone through typical editing workflows that include a mix of JPG compression, resizing, cropping, compositing, and rotation.

2. Applicable to Commercial and Open-source Solutions

Generative tooling is available in both commercial and open-source settings. Identification systems must provide cost-effective and OSS license compatible options to not encourage users of open-source generative solutions to bypass safeguards. Regulatory intervention may be required to prevent all Generative AI Systems (including open-source systems) from having safeguards disabled.

3. Decentralized to Respect Individual and Corporate Privacy

A centralized solution like Adobe's Content Credentials Cloud Registry introduces risk of oversharing that could pose existential threats to vulnerable populations and allow for

confidential corporate materials to leak to the public and competitors. A solution should not require any individual or company to disclose confidential information to competitors or the public.

4. Economical Despite Exponential Growth

In July 2023, only months after beta release, Adobe shared that Firefly had generated over 1 billion images. Midjourney has over 16 million users,[47]and Stable Diffusion boasts more than 10 million daily users "across all channels."[48] Any solution for identifying AI generated content must operate efficiently against trillions of pieces of content, and that content must remain identifiable well into the future as generative technologies continue to evolve. A viable solution must support rapid and accurate identification even as generative content grows to a multiple of all authentic content ever produced.

28.1. Who should be responsible for identifying a work as AI-generated?

Developers of Generative AI Models and AI Systems should be responsible for developing and integrating consistent and persistent identification systems into their tools and users should be required to use such integrations.

28.2. Are there technical or practical barriers to labeling or identification requirements?

Currently, there is no solution that consistently meets the criteria outlined in our response to Question 28.1. If a solution is possible, it will require the invention of an approach that is more indelible than anything available today, and that is available without burdensome licensing that would preclude open-source adoption.

It is our firm belief that the burden of identification, including the development of innovative identification techniques, should remain primarily with the providers of AI Models and AI Systems.

Below is a summary of the current inadequacies of leading technology solutions:

Embedded Metadata

Embedded metadata like IPTC DigitalSourceType may serve as a near term partial solution for compliance, but embedded metadata is nowhere close to indelible. Embedded metadata is frequently removed by screenshotting, passing content through publishing pipelines that optimize file size, or chat and social media platforms that strip metadata to protect their users from unintentionally leaking personally identifying information in their uploads.

Invisible Watermarks, Fingerprints, and Diffusion Noise Analysis

[47] As of October 28, 2023 Midjourney has 16,362,369 registered "members" on Discord. See https://discord.com/invite/midjourney for latest tally of users.
[48] See recent comments from Stability AI CEO, Emad Mostaque https://techcrunch.com/2022/10/17/stability-ai-the-startup-behind-stable-diffusion-raises-101m

Invisible watermarks are trivially defeated by bad actors and benign editing workflows like cropping and resizing regularly defeat even the most advanced invisible watermarks.[49] We are currently unaware of any available invisible watermark that will survive the type of edits that our customers regularly make while preparing content licensed from us for publication.

Analogous to invisible watermarks, fingerprints and diffusion noise analysis are methods to classify an image using innate patterns in the pixel data instead of a man-made invisible pattern. These techniques are just as vulnerable to disruption during routine editing workflows.

C2PA/CAI/Content Credentials/ProjectOrigin

Solutions that rely on a C2PA style chains of trust, like CAI and Project Origin, place a burden on human content creators to complicate their tool chains to keep track of changes made to a file in the ordinary course of preparing it for publication. These solutions require every consumer and professional workflow and system to adopt significant complexity in writing and validating C2PA metadata, which will undoubtedly delay adoption of safeguards for consumers by years.

C2PA can be used with embedded metadata, or store manifests externally in a blockchain or Content Credentials repository such as https://contentcredentials.org/verify. Given that embedded metadata is easily removed, externally hosted metadata will be required for C2PA to offer consistent verification. Externally hosted metadata brings with it the burden for companies and open-source providers to host expensive registries that need to scale to store trillions of images worth of metadata. To lookup content quickly and across publishers, registries will need to network with each other like DNS servers. Networked registries may lead to unacceptable sharing of data with competitors as the complete metadata for each file would be available from a public facing API.

28.3. If a notification or labeling requirement is adopted, what should be the consequences of the failure to label a particular work or the removal of a label?

We are reluctant to comment on potential consequences until specific labeling requirements have been proposed, as the appropriateness of any consequences for the failure to label a particular work or the removal of a label (e.g., fines or suspension of AI operating licenses granted by government) must take into account the nature of the requirements themselves. Therefore, we will reserve comment until there are specific proposals, except to say that any consequences must not in any way impact or threaten a rightsholder's ability to retain copyright protection in their works or otherwise run afoul of any international treaty obligations of the United States.

29. *What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?*

Please see the response to Question 28.2.

[49] See https://www.wired.com/story/artificial-intelligence-watermarking-issues/

Additional Questions About Issues Related to Copyright (Q. 30-34)

30. What legal rights, if any, currently apply to AI-generated material that features the name or likeness, including vocal likeness, of a particular person?

Many states have statutory law and/or common law regarding rights of publicity that prohibit the use of someone's image or likeness, including voice, for commercial purposes regardless of whether the likeness is an actual depiction of the individual or an AI-created depiction. However, these laws do not prevent editorial or expressive uses protected by the First Amendment, such as uses in a parody or in the context of another work such as a documentary or film.

31. Should Congress establish a new federal right, similar to state law rights of publicity, that would apply to AI-generated material? If so, should it preempt state laws or set a ceiling or floor for state law protections? What should be the contours of such a right?

A new federal right of publicity that would apply to AI generated material could be helpful to protect individuals from the potential harms posed by deep fakes and other unauthorized content. However, until the details of such a right have been proposed, it is premature to say if similar state laws should be preempted. Regardless, any new federal right of publicity should be carefully considered so that constitutionally protected expression is not unduly limited. Accordingly, legislation should include explicit exemptions for First Amendment-protected expression. For example, New York State's right of publicity law includes a provision that exempts good-faith sales and licensing of visual works that do not authorize the user to violate the law.[50] Any new federal right should include similar exemptions that do not limit the market for legitimate licensing.

32. Are there or should there be protections against an AI system generating outputs that imitate the artistic style of a human creator (such as an AI system producing visual works "in the style of" a specific artist)? Who should be eligible for such protection? What form should it take?

No response.

33. With respect to sound recordings, how does section 114(b) of the Copyright Act relate to state law, such as state right of publicity laws? Does this issue require legislative attention in the context of generative AI?

No response

34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study *No response*

[50] See New York Civil Rights Law Chapter 6, Article 5 Section 50-F Right of publicity. Subdivision 10 reads, *"Nothing in this section shall apply to a person that offers a service that displays, offers for sale or license, sells or licenses a work of art or other visual work, or audiovisual work, to a user, provided the terms of such sale or license do not authorize such user to engage in acts that constitute a violation of this section." See* https://www.nysenate.gov/legislation/laws/CVR/50-F *for full text of Section 50-F*

Appendix A

Amended Complaint, Getty Images (US), Inc. v. Stability AI, Inc. and Stability AI, Ltd.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

GETTY IMAGES (US), INC. Plaintiff, v. STABILITY AI, LTD. and STABILITY AI, INC. Defendants.	)	C.A. No. 23-135 (GBW) **DEMAND FOR JURY TRIAL**

AMENDED COMPLAINT

Plaintiff Getty Images (US), Inc. ("Getty Images" or "Plaintiff"), by and through its undersigned attorneys, for its Amended Complaint against Defendants Stability AI, Ltd. and Stability AI, Inc. (collectively "Stability AI" or "Defendants"), hereby alleges as follows:

NATURE OF ACTION

1. This case arises from Stability AI's brazen infringement of Getty Images' intellectual property on a staggering scale. Upon information and belief, Stability AI has copied *more than 12 million* photographs from Getty Images' collection, along with the associated captions and metadata, without permission from or compensation to Getty Images, as part of its efforts to build a competing business. As part of its unlawful scheme, Stability AI has removed or altered Getty Images' copyright management information, provided false copyright management information, and infringed Getty Images' famous trademarks.

2. Getty Images brings this action to recover damages that it has suffered and is continuing to suffer, and to prevent the irreparable harm caused by Stability AI's intentional and willful acts in violation of United States and Delaware law.

3. Getty Images is one of the world's leading creators and distributors of digital content. At great expense, over the course of nearly three decades, Getty Images has curated a

collection of hundreds of millions of premium quality visual assets, most of which are still, photographic images. Many of these images were created by Getty Images staff photographers as works made-for-hire, others have been acquired by Getty Images from third parties with an assignment of the associated copyrights, and the remainder have been licensed to Getty Images by its hundreds of content partners or hundreds of thousands of contributing photographers, who rely on the licensing income Getty Images generates for them.

4. Getty Images makes hundreds of millions of visual assets available to customers throughout the world and in this District via websites, including but not limited to www.gettyimages.com and www.istock.com. The visual assets on Getty Images' websites are accompanied by: (i) titles and captions which are themselves original and creative copyrighted expression; (ii) watermarks with credit information and content identifiers that are designed to deter infringing uses of the content; and (iii) metadata containing other copyright management information.

5. Getty Images serves creative, corporate, and media customers in more than 200 countries around the world, and its imagery helps its customers produce work which appears every day in the world's most influential newspapers, magazines, advertising campaigns, films, television programs, books and websites. In appropriate circumstances, and with safeguards for the rights and interests of its photographers and contributors and the subjects of the images in its collection, Getty Images also licenses the use of its visual assets and associated metadata in connection with the development of artificial intelligence and machine learning tools. Getty Images has licensed millions of suitable digital assets to leading technology innovators for a variety of purposes related to artificial intelligence and machine learning.

6. Getty Images' visual assets are highly desirable for use in connection with artificial intelligence and machine learning because of their high quality, and because they are accompanied by content-specific, detailed captions and rich metadata.

7. Upon information and belief, Stability AI was founded in 2020 by Emad Mostaque, a former hedge fund executive, as a for-profit company. According to press reports in October 2022, Stability AI raised more than $100 million from venture capital investors and was already valued at $1 billion. According to more recent press reports, Stability AI is now seeking to raise even more money at a valuation of approximately $4 billion. On the back of intellectual property owned by Getty Images and other copyright holders, Stability AI has created an image-generating model called Stable Diffusion that uses artificial intelligence to deliver computer-synthesized images in response to text prompts. In additional to offering open-source versions of Stable Diffusion, Stability AI offers a revenue-generating user interface called DreamStudio that is powered by its Stable Diffusion model. DreamStudio enables users to obtain images from the Stable Diffusion model on their own personal computers without the need for software installation or coding knowledge, and Stability AI charges fees for that service.

8. Rather than attempt to negotiate a license with Getty Images for the use of its content, and even though the terms of use of Getty Images' websites expressly prohibit unauthorized reproduction of content for commercial purposes such as those undertaken by Stability AI, Stability AI copied at least 12 million copyrighted images from Getty Images' websites, along with associated text and metadata, in order to train its Stable Diffusion model.

9. Stability AI now competes directly with Getty Images by marketing Stable Diffusion and its DreamStudio interface to those seeking creative imagery, and its infringement of Getty Images' content on a massive scale has been instrumental to its success to date.

10. Upon information and belief, Stability AI was well aware that the content it was scraping without permission from Getty Images' websites was protected by U.S. copyright law.

11. Often, the output generated by Stable Diffusion has contained a modified version of a Getty Images watermark, creating confusion as to the source of the images and falsely implying an association with Getty Images. While some of the output generated through the use of Stable Diffusion is aesthetically pleasing, other output is of much lower quality and at times ranges from the bizarre to the grotesque. Stability AI's incorporation of Getty Images' marks into low quality, unappealing, or offensive images dilutes those marks in further violation of federal and state trademark laws.

12. Getty Images therefore brings this action alleging claims under the Copyright Act of 1976, 17 U.S.C. §101 *et seq.*, the Lanham Act, 15 U.S.C. § 1051 *et seq.*, and Delaware trademark and unfair competition laws to bring an end to Stability AI's blatantly infringing conduct in the United States and in Delaware and to obtain redress for Stability AI's callous disregard for its intellectual property rights.

PARTIES

13. Plaintiff Getty Images (US), Inc. is a New York corporation with headquarters in Seattle, Washington. It is the owner or exclusive licensee of the copyrights subject to the copyright infringement claims at issue and the owner of the trademarks at issue.

14. Upon information and belief, Defendant Stability AI, Inc. is a Delaware corporation with headquarters in London, UK.

15. Upon information and belief, Defendant Stability AI, Ltd. is a UK corporation with headquarters in London, UK. As set forth more fully below, Defendants Stability AI, Ltd. and Stability AI, Inc. are alter egos of one another and operate as a single enterprise.

JURISDICTION AND VENUE

16. This action arises under the Copyright Act of 1976, 17 U.S.C. §101 *et seq.*, the Lanham Act, 15 U.S.C. § 1051 *et seq.*, and Delaware trademark and unfair competition laws. This Court has subject matter jurisdiction pursuant to 28 U.S.C. §§ 1331, 1338, and 1367.

17. This Court has personal jurisdiction over Defendant Stability AI, Inc. because Stability AI, Inc. is incorporated in Delaware.

18. Defendant Stability AI, Ltd. is an alter ego of and operates as a single enterprise with Defendant Stability AI, Inc. The two corporations share the same CEO and founder: Mr. Mostaque. Upon information and belief, in addition to serving as CEO and Director of Stability AI, Inc., Mr. Mostaque controls 75% or more of the voting rights, 75% or more of the shares, and has the right to appoint or remove a majority of the board of directors of Stability AI, Ltd. Stability AI, Inc. and Stability AI, Ltd. also present themselves as a single enterprise: their principal offices are located at the same physical London address and share both an email domain (@stability.ai) and website (https://stability.ai/).

19. According to Dun & Bradstreet, Stability AI, Ltd. is a subsidiary of Stability AI, Inc. and, as of November 2022, Stability AI, Ltd.'s sole share was owned by Stability AI, Inc. And, according to the records of the Delaware Secretary of State, Stability AI, Inc.'s corporate charter was voided for non-payment of taxes and/or failure to file a complete annual report in 2022, and Stability AI, Inc. subsequently filed a certificate to revive its charter, indicating that Stability AI, Inc. is not an independently-operating company.

20. Upon information and belief, Stability AI, Ltd. employs all of the company's employees and conducts all of the company's activities, while Stability AI, Inc. is a shell holding company, which has no employees or day-to-day operations. But, according to SEC filings, in

October 2022, Stability AI, Inc. raised over $75 million through a securities offering (including around $11 million of convertible indebtedness and accrued interest). Upon information and belief, the funds raised by Stability, AI, Inc. are used to fund the activities of Stability AI, Ltd., including those described in this Amended Complaint.

21. This Court also has personal jurisdiction over Defendant Stability AI, Ltd. based on Stability AI, Ltd.'s contacts with Delaware and the United States.

22. Stability AI, Ltd. operates a website that is accessible to internet users in Delaware and elsewhere in the United States. From that website, users throughout the United States, including in Delaware, can access Stability AI's offerings, such as Stable Diffusion and DreamStudio.

23. The Stability AI, Ltd. website does not specifically target users in any one state of the United States, and instead targets users across the United States, including users located in Delaware.

24. Upon information and belief, Stability AI, Ltd. maintains cloud computing and physical server resources in the United States.

25. The Stability AI, Ltd. website expressly states that the site (i.e., https://stability.ai/) and its content are "protected by copyright, trade dress, trademark, moral rights, and other intellectual property laws in the United States, the United Kingdom, and other international jurisdictions." As a result, Stability AI, Ltd. has demonstrated its intent to avail itself of jurisdiction and the legal protections of the United States.

26. Accordingly, Stability AI, Ltd. has sufficient contacts with the United States to be subject to personal jurisdiction in Delaware. *See* Fed. R. Civ. P. 4(k)(2).

27. Venue is proper in this District pursuant to 28 U.S.C. § 1391(b) because Defendant Stability AI, Inc. is subject to personal jurisdiction in this District. Venue is also proper in this District pursuant to 28 U.S.C. § 1400(a), because Stability AI or its agents reside or may be found in this District.

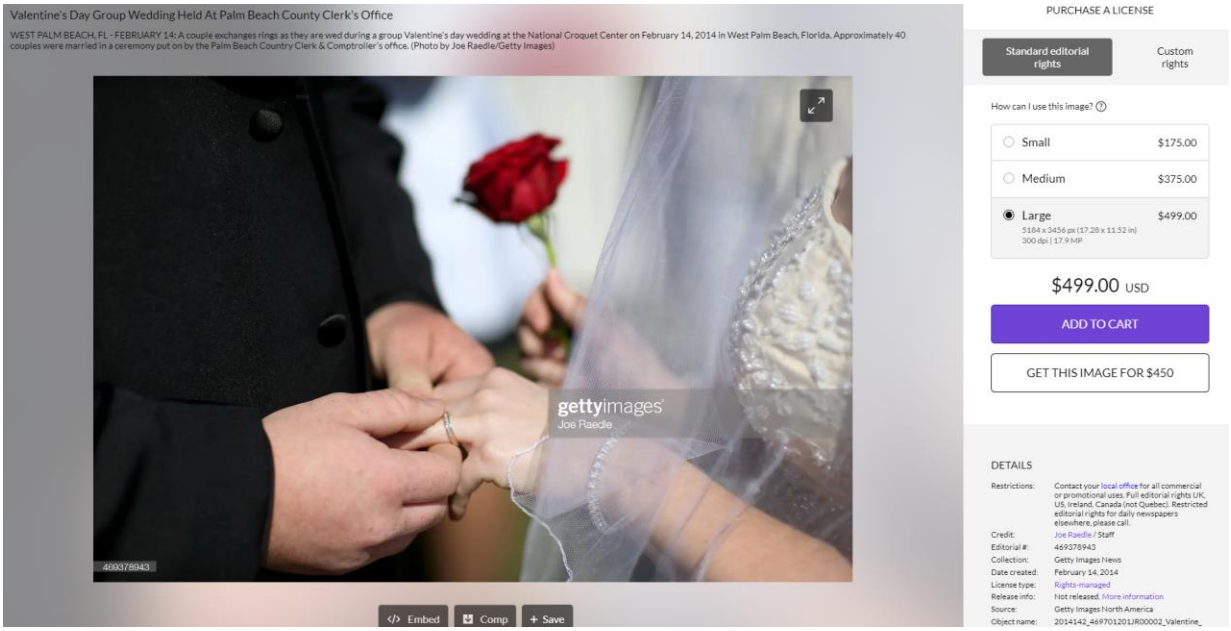
ALLEGATIONS COMMON TO ALL CLAIMS FOR RELIEF

A. Getty Images, Its Extensive Collection, and Its Worldwide Reputation for Premium Visual Content

28. Getty Images is a preeminent global visual content creator and a leading source for visual content around the world. Getty Images operates websites for the purpose of licensing its works, including, *inter alia*, at www.gettyimages.com and www.istock.com. Its collection, which currently contains hundreds of millions of visual assets, is renowned worldwide for its unmatched depth, breadth, and quality. That visual content is included in a robust database (the "Database") that also contains detailed, original text titles and captions associated with the individual photographs and rich, image-specific metadata to provide the highest quality user experience to customers and to ensure appropriate compensation for contributors and content partners.

29. By visiting Getty Images' websites, its customers and potential customers can search and browse its collection before purchasing a license for specific content. For example, customers looking for an image from a wedding might search "a couple exchanges rings." Among the search results, they might find the following image available for license with an accompanying title that reads, "Valentine's Day Group Wedding Held at Palm Beach County Clerk's Office," a caption that reads, "A couple exchanges rings as they are wed during a group

Valentine's day wedding at the National Croquet Center on February 14, 2014 in West Palm Beach, Florida" and a photo credit that reads "(Photo by Joe Readle/Getty Images)":[1]



30. As the foregoing example reflects, the search results contain, in addition to images responsive to the search terms, watermarks on the images to deter infringing uses, credits and other metadata, and options for purchasing a license for further use.

31. Getty Images has more than 500,000 contributors (80,000 of which are exclusive to Getty Images), over 300 premium content partners, more than 115 staff photographers, videographers, and other content experts who guide and contribute to the creation of award-winning content, and a unique and comprehensive visual archive collection covering a broad range of subject matter. Contributors choose to work with Getty Images to benefit from its reputation and goodwill as a preeminent content licensor, its robust platform, its global distribution network, and the royalty income Getty Images generates for them.

[1] https://www.gettyimages.com/detail/news-photo/couple-exchanges-rings-as-they-are-wed-during-a-group-news-photo/469378943?phrase=a%20couple%20exchanges%20rings&adppopup=true.

32. Getty Images' customers come to Getty Images for its easy-to-use platform, its comprehensive suite of content (including certain types of content for which authorized copies are exclusive to Getty Images), its variety of licensing options and services, and the assurance that the images they obtain from Getty Images will not infringe third-party copyrights.

B. Getty Images' Intellectual Property Rights and Terms of Use

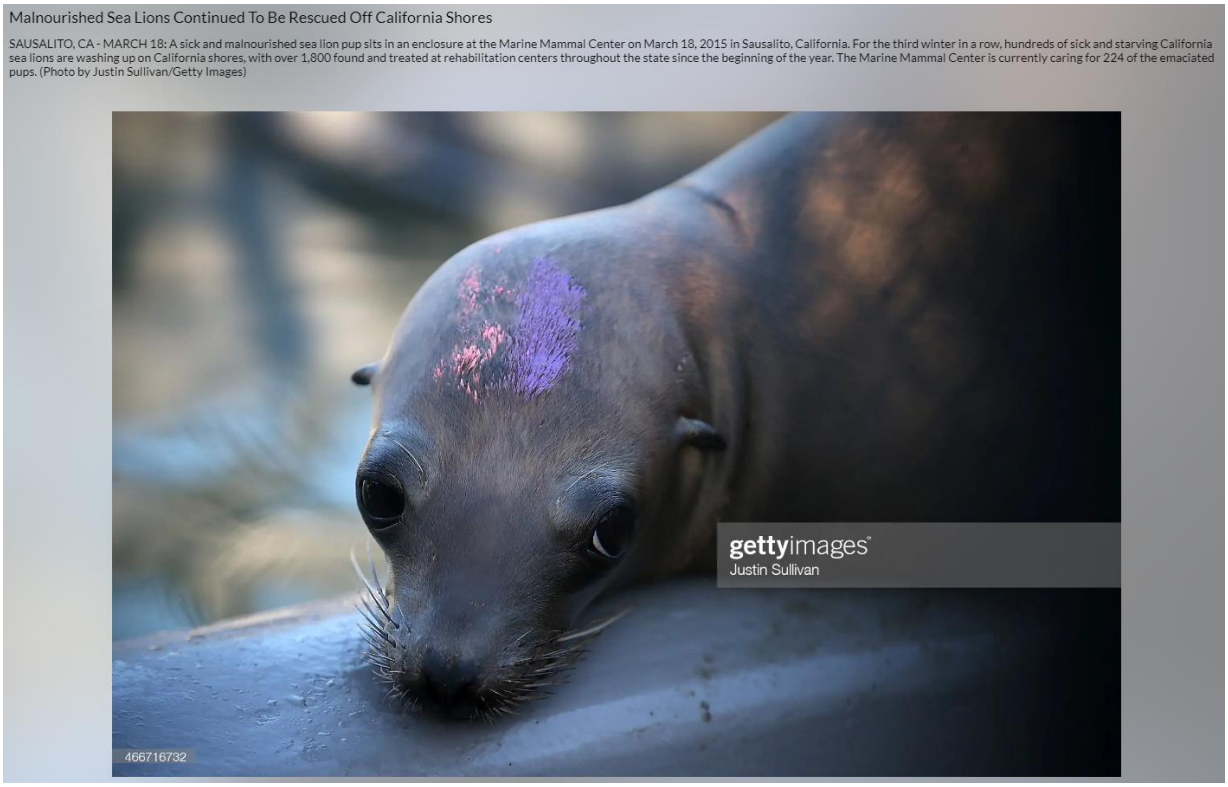
1. Copyright

33. Most of the images and videos displayed on Getty Images' websites are original, creative works that enjoy protection under U.S. copyright laws. For many of these visual assets, including all of the assets subject to the copyright infringement claims at issue in this action, Getty Images either owns the copyright or is an exclusive licensee; for others, Getty Images is a non-exclusive licensee.

34. For purposes of the copyright infringement claims set forth herein and establishing the unlawful nature of Stability AI's conduct, Getty Images has selected 7,216 examples from the millions of images that Stability AI copied without permission and used to train one or more versions of Stable Diffusion. The copyrights for each of these images (as well as for many other images) have been registered with the U.S. Copyright Office. A list of these works, together with their copyright registration numbers, is attached as Exhibit A.

35. As noted above, for the images displayed on its websites, Getty Images also typically provides a detailed corresponding title and caption. Image titles and captions, which are authored either by a Getty Images staff member or by an image contributor or partner, typically reflect originality and creative choices. For example, for the image below, the accompanying title reads, "Malnourished Sea Lions Continued To Be Rescued Off California Shores" and the accompanying caption reads: "A sick and malnourished sea lion pup sits in an

enclosure at the Marine Mammal Center on March 18, 2015 in Sausalito, California. For the third winter in a row, hundreds of sick and starving California sea lions are washing up on California shores, with over 1,800 found and treated at rehabilitation centers throughout the state since the beginning of the year. The Marine Mammal Center is currently caring for 224 of the emaciated pups."[2]



36. Each of the images available through Getty Images' websites has an associated page that contains a unique URL pointing to a location where the image is stored together with an "alt text" tag containing the image title and caption. The image URLs, titles, and captions, along with other current metadata for each image, such as keywords and author and ownership data, are populated from the Database.

[2] https://www.gettyimages.com/detail/news-photo/sick-and-malnourished-sea-lion-pup-sits-in-an-enclosure-at-news-photo/466716732

37. Getty Images has spent years coordinating and arranging the Database, including, *inter alia*, by setting criteria for inclusion of images, selecting specific images for inclusion, creating and incorporating detailed captions and other text paired with images, creating and assigning unique asset identifiers that can be linked to specific contributors, and arranging the contents of the Database so that the Database is searchable and results can be filtered. Additionally, Getty Images has and continues to invest significantly in maintaining the contents of the Database. Between 2017 and 2020 alone, Getty Images and its affiliates invested more than $200 million to maintain the Database.

38. Getty Images has registered its copyright of the Database with the United States Copyright Office. The copyright registration number is TXu002346096.

2. Trademarks and Goodwill

39. Getty Images' name and trademarks are renowned in the U.S. and around the world. Customers perform over 2.7 billion searches annually on the Getty Images' websites, which exist in 23 languages. Through its full range of content solutions, Getty Images served over 836,000 purchasing customers in the last year alone, with customers from almost every country in the world, ranging from media outlets, advertising agencies, and corporations of all sizes to individual creators. Customers rely on Getty Images for the best content and service, and trust the trademarks and service marks associated with its content.

40. Since its founding in 1995, Getty Images has been using its name and associated trademarks in commerce continuously in connection with the distribution, promotion, and marketing of its services and visual content in the United States, including the uses described above. Getty Images has used its name and trademarks exclusively and extensively in the United

States and in Delaware, and its trademarks are widely recognized as representing premium quality visual content.

41. Getty Images uses its name and trademarks prominently on the Getty Images websites. Each image available for viewing and purchase prominently displays a watermark that contains an affiliated trademark, as illustrated in the images depicted in paragraphs 29 and 35 above.

42. Getty Images owns trademarks registered on the Principal Register in the United States Patent & Trademark Office ("USPTO") relating to its iconic brand. True and correct copies of the federal registration certificates evidencing Getty Images' ownership of the trademarks shown below are attached hereto as Exhibit B.

Mark Name	Reg. Number	Reg. Date
GETTY IMAGES	2,656,652	12/03/2002
GETTY IMAGES	2,837,208	04/27/2004
GETTY IMAGES	2,842,851	05/18/2004
GETTY IMAGES	2,844,647	05/25/2004
GETTY IMAGES	3,603,335	04/07/2009
GETTY IMAGES	4,968,996	05/31/2016
GETTY IMAGES	4,968,997	05/31/2016
GETTY IMAGES	5,200,414	05/09/2017

43. Getty Images also owns common law rights in the mark GETTY IMAGES. Together with Getty Images' federally registered trademarks, these are referred to collectively as the "Getty Images Marks."

3. Website Terms and Conditions

44. Stability AI accessed Getty Images' collection of visual assets through Getty Images' public-facing websites. The Getty Images websites from which Stability AI copied images without permission are subject to express terms and conditions of use which, among other things, expressly prohibit, *inter alia*: (i) downloading, copying or re-transmitting any or all

of the website or its contents without a license; and (ii) using any data mining, robots or similar data gathering or extraction methods. Such restrictions apply not only to the photographic images and videos that Getty Images licenses, but also to the valuable and proprietary title and caption information, keywords, and other metadata associated with the visual assets, all of which is highly desirable for use in connection with developing AI tools such as Stable Diffusion.

C. Stability AI Infringes Getty Images' Copyrights on an Enormous Scale and Exploits Getty Images' Resources for its Commercial Benefit

45. Upon information and belief, Stability AI was founded in 2020 and is engaged in the development of tools and models to generate digital content using artificial intelligence.

46. Stability AI created and maintains a model called Stable Diffusion. Upon information and belief, Stability AI utilizes the following steps from input to output:

a. First, Stability AI copies billions of text-and-image pairings—like those available on Getty Images' websites—and loads them into computer memory to train a model.

b. Second, Stability AI encodes the images, which involves creating smaller versions of the images that take up less memory. Separately, Stability AI also encodes the paired text. Stability AI retains and stores copies of the encoded images and text as an essential element of training the model.

c. Third, Stability AI adds visual "noise" to the encoded images, *i.e.*, it further alters the images so that it is incrementally harder to discern what is visually represented because the images have been intentionally degraded in visual quality in order to "train" the model to remove the "noise." By intentionally adding visual noise to the existing images with associated text, Stability AI teaches the model to generate output images to be consistent with a particular text description

(*e.g.*, "a dog playing on the beach during sunset").

d. Fourth, the model decodes the altered image and teaches itself to remove the noise by comparing the decoded image to the original image and text descriptions that have been copied and stored. By learning to decode noise, the model learns to deliver images similar to—and, in some cases, substantially similar to—the original without noise.

47. Upon information and belief, the third and fourth steps described in the preceding paragraph are part of "training" the model to allow Stable Diffusion to understand the relationships between text and associated images and to use that knowledge to computationally produce images in response to text prompts, as explained further below.

48. Stable Diffusion was trained on 5 billion image-text pairs from datasets prepared by non-party LAION, a German entity that works in conjunction with and is sponsored by Stability AI. Upon information and belief, Stability AI provided LAION with both funding and significant computing resources to produce its datasets in furtherance of Stability AI's infringing scheme.

49. Upon information and belief, LAION created the datasets of image-text pairs used by Stability AI by scraping links to billions of pieces of content from various websites, including Getty Images' websites.

50. Upon information and belief, Stability AI followed links included in LAION's dataset to access specific pages on Getty Images' websites and copied many millions of copyrighted images and associated text. Such copying was done without Getty Images' authorization and in violation of the express prohibitions against such conduct contained in its websites' terms of use.

51. Upon information and belief, Stability AI then created another copy of the content to encode it into a form its model could interpret.

52. Upon information and belief, Stability AI then created yet additional copies with visual noise added, while retaining encoded copies of the original images without noise for comparison to help train its model.

53. Upon information and belief, the unauthorized copies of Getty Images' content made by Stability AI are neither transitory nor ephemeral, and they were made with the express aim of enabling Stability AI to supplant Getty Images as a source of creative visual imagery.

54. To date, Getty Images has identified over 12 million links to images and their associated text and metadata on its websites contained in the LAION datasets that were used to train Stable Diffusion. Among the millions of links was a link to the photograph of the couple exchanging rings displayed in paragraph 29 above as well as to each of the other images identified in Exhibit A.

55. Getty Images' content is extremely valuable to the datasets used to train Stable Diffusion. Getty Images' websites provide access to millions of high quality images and a vast array of subject matter. High quality images such as those offered by Getty Images on its websites are more useful for training an AI model such as Stable Diffusion than low quality images because they contain more detail or data about the image that can be copied. By contrast, a low quality image, such as one that has been compressed and posted as a small thumbnail on a typical social media site, is less valuable because it only provides a rough, poor quality framework of the underlying image and may not be accompanied by detailed text or other useful metadata.

56. Stability AI has developed and released different versions of Stable Diffusion over time, including, upon information and belief, to users located in Delaware. The core dataset used to train Stable Diffusion version 2 was a subset of LAION 5B called LAION-Aesthetics,[3] which was created to exclude images that were not sufficiently aesthetically pleasing.[4] Targeting its copying in this way allowed Stability AI to further benefit from Getty Images' efforts over many years to amass its renowned collection of high quality images and from the significant investments required to generate such a collection and to develop and maintain the Database in which it is stored.

57. Second, Getty Images' websites include both the images and corresponding detailed titles and captions and other metadata. Upon information and belief, the pairings of detailed text and images has been critical to successfully training the Stable Diffusion model to deliver relevant output in response to text prompts. If, for example, Stability AI ingested an image of a beach that was labeled "forest" and used that image-text pairing to train the model, the model would learn inaccurate information and be far less effective at generating desirable outputs in response to text prompts by Stability AI's customers. Furthermore, in training the Stable Diffusion model, Stability AI has benefitted from Getty Images' image-text pairs that are not only accurate, but detailed. For example, if Stability AI ingested a picture of Lake Oroville in California during a severe drought with a corresponding caption limited to just the word "lake," it would learn that the image is of a lake, but not which lake or that the photograph was taken during a severe drought. If a Stable Diffusion user then entered a prompt for "California's Lake Oroville during a severe drought" the output image might still be one of a lake, but it would

[3] https://stability.ai/blog/stable-diffusion-announcement.

[4] https://laion.ai/projects/.

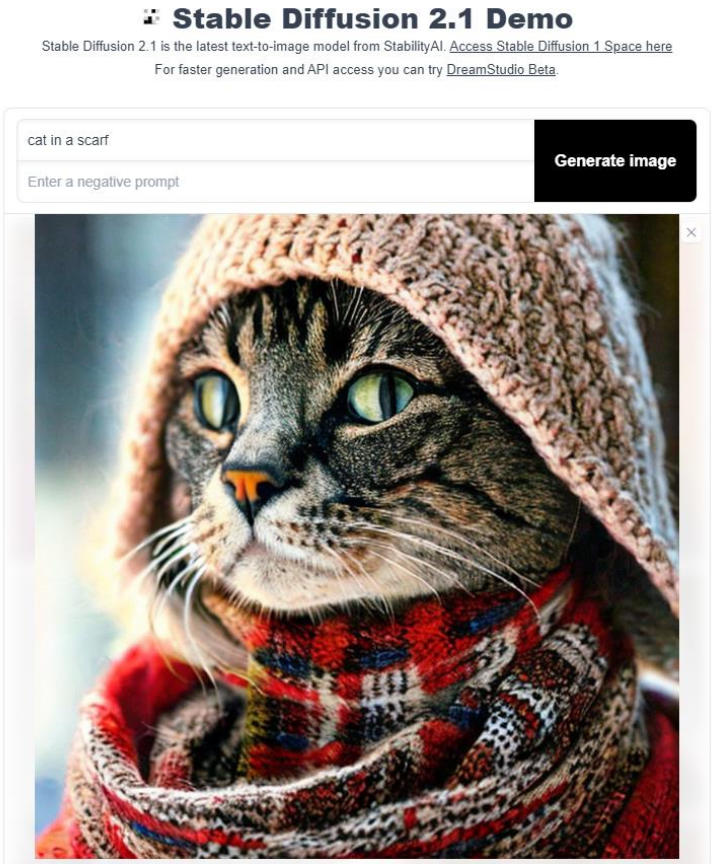
be much less likely to be an image of Lake Oroville during a severe drought because the synthesis engine would not have the same level of control that allows it to deliver detailed and specific images in response to text prompts.

58. Upon information and belief, when Stability AI ingested the image below of Lake Oroville with a corresponding caption that reads "A section of Lake Oroville is seen nearly dry on August 19, 2014 in Oroville, California. As the severe drought in California continues for a third straight year, water levels in the State's lakes and reservoirs is reaching historic lows. Lake Oroville is currently at 32 percent of its total 3,537,577 acre feet,"[5] its use of the accompanying text enabled the model to learn even more about the image and its contents and thus generate output that competes with Getty Images' own offerings much more effectively.

[5] https://www.gettyimages.com/detail/news-photo/section-of-lake-oroville-is-seen-nearly-dry-on-august-19-news-photo/453834006

D. Stability AI Competes Commercially with Getty Images

59. Once an artificial intelligence model like Stable Diffusion has been trained on enough data to learn the relationship between text prompts and images, it can be used to generate new images derived from the images and text the model's creator has copied. For example, if a model has been trained with image-text pairs of cats and image-text pairs of clothing, then a user can use the text prompt "cat in a scarf" and the model will generate an image that looks like a cat in a scarf:



60. To be clear, the image above is not a photograph of an actual cat wearing an actual scarf. It is a computer-synthesized image that *resembles* a cat wearing a scarf. Upon information and belief, Stability AI was able to generate the image above because it used enough images of real cats paired with rich text captions and images of real scarves with rich text

captions to train Stable Diffusion that the model can generate this type of output. Stable Diffusion is able to combine what it has learned to generate this artificial image, but only because it was trained on proprietary content belonging to Getty Images and others.

61. As a result, Stable Diffusion at times produces images that are highly similar to and derivative of the Getty Images proprietary content that Stability AI copied extensively in the course of training the model. Indeed, independent researchers have observed that Stable Diffusion sometimes memorizes and regenerates specific images that were used to train the model.[6]

62. In many cases, and as discussed further below, the output delivered by Stability AI includes a modified version of a Getty Images watermark, underscoring the clear link between the copyrighted images that Stability AI copied without permission and the output its model delivers. In the following example, the image on the left is another original, watermarked image copied by Stability AI and used to train its model and the watermarked image on the right is output delivered using the model:

[6] *See, e.g.*, Nicholas Carlini et al., Extracting Training Data from Diffusion Models (2023), https://arxiv.org/pdf/2301.13188.pdf; *see also* Gowthami Somepalli et al., Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models (2022), https://arxiv.org/pdf/2212.03860.pdf.

63. Upon information and belief, Stability AI offers Stable Diffusion as open source software, meaning that Stability AI permits third party developers to access, use, and further develop the model without paying license fees to Stability AI. Those third parties benefit from Stability AI's infringement of Getty Images' copyrights and, in turn, Stability AI benefits from the widespread adoption of its model.

64. While Stability AI has made Stable Diffusion open source, Stability AI is also directly monetizing the tool through a commercial platform it calls DreamStudio. DreamStudio allows customers to access Stable Diffusion to generate images without the need for any of their own heavy-duty processing power, software installation, or coding knowhow. According to Mr. Mostaque, Stability AI plans to further monetize Stable Diffusion by training and deploying customized, non-open source versions of Stable Diffusion for customers for use on a large scale,

and Stability AI reportedly was valued at $1 billion by late 2022 and is seeking additional funding at a valuation of approximately $4 billion.[7]

65. Upon information and belief, although Stability AI only released DreamStudio in August 2022, millions of people already have used DreamStudio and collectively created hundreds of millions of images. Yet Stability AI has not paid a cent to Getty Images or other content owners from which it reproduced copyrighted content without permission to train its highly lucrative model.

66. The gravity of Stability AI's brazen theft and freeriding is compounded by the fact that, by utilizing Getty Images' copyrighted content for artificial intelligence and machine learning, Stability AI is stealing a service that Getty Images already provides to paying customers in the marketplace for that very purpose. Getty Images has licensed millions of suitable digital assets for a variety of purposes related to artificial intelligence and machine learning in a manner that respects personal and intellectual property rights. While Getty Images licenses its proprietary content to responsible actors in appropriate circumstances, Stability AI has taken that same content from Getty Images without permission, depriving Getty Images and its contributors of fair compensation, and without providing adequate protections for the privacy and dignity interests of individuals depicted.

E. Stability AI's Attempts to Circumvent Getty Images' Watermarks

67. As noted in paragraph 41 above, each copyrighted image on Getty Images' public-facing websites contains a watermark that is intended to indicate provenance and prevent

[7] https://techcrunch.com/2022/10/17/stability-ai-the-startup-behind-stable-diffusion-raises-101m/; https://fortune.com/2023/03/04/stability-ai-raise-funds-4-billion-valuation-artificial-intelligence-captivates-investors/.

infringement. The watermark includes both a Getty Images-owned mark and credit information for the image.

68. Upon information and belief, Stability AI has knowingly removed Getty Images' watermarks from some images in the course of its copying as part of its infringing scheme. At the same time, however, as discussed above, the Stable Diffusion model frequently generates output bearing a modified version of the Getty Images watermark, even when that output is not bona fide Getty Images' content and is well below Getty Images' quality standards. Examples of this practice include:

69. Making matters worse, Stability AI has caused the Stable Diffusion model to incorporate a modified version of the Getty Images' watermark to bizarre or grotesque synthetic imagery that tarnishes Getty Images' hard-earned reputation, such as the image below:

70. Upon information and belief, Stability AI is well aware that Stable Diffusion generates images that include distorted versions of Getty Images' watermark and other watermarks, but it has not modified its model to prevent that from happening.

71. Upon information and belief, unless enjoined by this Court, Stability AI intends to continue to infringe upon Getty Images' copyrights and trademarks in the United States and otherwise to profit from its unauthorized use of Getty Images' intellectual property. Getty Images has no adequate remedy at law to redress all of the injuries that Stability AI has caused, and intends to continue to cause, by its conduct. Getty Images will continue to suffer irreparable harm until Stability AI's infringing conduct is enjoined by this Court.

CLAIM I

Copyright Infringement (17 U.S.C. § 101 *et seq.*)

72. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 71 above.

73. Getty Images is the owner or exclusive licensee of copyrights identified in Exhibit A, and therefore is entitled to the exclusive rights under copyright law associated therewith, including the rights set forth in 17 U.S.C § 106.

74. Getty Images has obtained copyright registrations in the United States for each of the works identified in Exhibit A.

75. Getty Images is the owner of, and has obtained a U.S. copyright registration for, the Database.

76. Stability AI obtained access to the registered images and the associated titles, captions, and other metadata in the Database through Getty Images' websites.

77. By and through the actions alleged above, Stability AI has infringed and will continue to infringe Getty Images' copyrights in the United States by, *inter alia*, reproducing Getty Images' copyrighted works and creating derivative works therefrom without any authorization from Getty Images.

78. Stability AI's acts of copyright infringement have been intentional, willful, and in callous disregard of Getty Images' rights. Stability AI knew at all relevant times that the content on Getty Images' websites is copyrighted, that Getty Images is in the business of licensing visual content, and that its acts were in violation of the terms of use of Getty Images' websites.

79. Stability AI engaged in the infringing acts described herein for its own commercial benefit.

80. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further acts of infringement and continuing to use and distribute Stable Diffusion models trained using Getty Images' copyrighted content without permission, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from engaging in any further infringement of Getty Images' content.

81. Getty Images is further entitled to recover from Stability AI the damages it has sustained and will sustain as a result of the infringing acts alleged above, together with any additional profits obtained by Stability AI. The amount of such damages and profits cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

82. For any infringing acts in the United States occurring after registration of the applicable Getty Images' copyrights, Getty Images is entitled, at its election, as an alternative to an award of actual damages and any additional profits earned by Stability AI, to recover statutory damages of up to $150,000 for each infringed work.

83. Getty Images is entitled to recover its full costs in prosecuting its copyright infringement claims in this action and its attorneys' fees.

CLAIM II

Providing False Copyright Management Information in Violation of 17 U.S.C. § 1202(a)

84. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 83 above.

85. The watermarks that Getty Images applies to images made available on its public-facing websites constitute copyright management information for purposes of Section 1202 of the Copyright Act, 17 U.S.C. § 1202.

86. By applying a modified version of Getty Images' watermarks to output generated through use of Stable Diffusion and the DreamStudio interface, Stability AI has provided false copyright management information in violation of 17 U.S.C. § 1202(a). Stability AI's provision of false copyright management information has been done knowingly and with the intent to induce, enable, facilitate, or conceal infringement of Getty Images' copyrights.

87. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further acts of providing false copyright management information, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability

AI, its agents, affiliates, employees and all persons acting in concert with it from providing false copyright management information.

88. Getty Images is further entitled to recover from Stability AI the damages it has sustained and will sustain as a result of the unlawful acts alleged above, together with any additional profits obtained by Stability AI. The amount of such damages and profits cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

89. Getty Images is entitled, at its election, as an alternative to an award of actual damages and any additional profits earned by Stability AI, to recover statutory damages of up to $25,000 for each violation of Section 1202(a).

90. Getty Images is entitled to recover its full costs and attorneys' fees in prosecuting its claims under Section 1202(a).

CLAIM III

Removal or Alteration of Copyright Management Information in Violation of Section 1202(b)

91. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 90 above.

92. Stability AI has intentionally removed or altered Getty Images' watermarks and metadata associated with the images Stability AI impermissibly copied from Getty Images' websites. Such watermarks and metadata contain copyright management information. Stability AI's removal or alteration of Getty Images' copyright management information has been done knowingly and with the intent to induce, enable, facilitate, or conceal infringement of Getty Images' copyrights.

93. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination

and, unless permanently enjoined from further acts of removing or altering copyright management information, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from removing or altering Getty Images' copyright management information.

94. Getty Images is further entitled to recover from Stability AI the damages it has sustained and will sustain as a result of the unlawful acts alleged above, together with any additional profits obtained by Stability AI. The amount of such damages and profits cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

95. Getty Images is entitled, at its election, as an alternative to an award of actual damages and any additional profits earned by Stability AI, to recover statutory damages of up to $25,000 for each violation of Section 1202(b).

96. Getty Images is entitled to recover its full costs and attorneys' fees in prosecuting its claims under Section 1202(b).

CLAIM IV

Trademark Infringement in Violation of Section 32 of the Lanham Act, 15 U.S.C. § 1114(1)

97. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 96 above.

98. Getty Images has expended substantial time, money, and resources collecting, distributing, promoting, marketing, and advertising the millions of images it offers on its websites and the Getty Images Marks associated therewith.

99. The Getty Images Marks are in full force and effect. Getty Images has never abandoned them, nor has Getty Images ever abandoned the goodwill of its businesses in

connection thereto. For example, Getty Images continues to use and prominently display Getty Images Marks on its websites, as well as on and in connection with the many millions of images it offers. Getty Images intends to continue to preserve and maintain its rights with respect to the Getty Images Marks.

100. The Getty Images Marks are distinctive and have become associated in the minds of the public with Getty Images, its brand, and its reputation for high-quality visual content.

101. The Getty Images Marks and the goodwill of the business associated with them in the United States are of great and significant value to Getty Images.

102. Getty Images' use of the Getty Images Marks and Stability AI's infringing uses of the same marks are in competitive proximity to one another, as they are both used in connection with, *inter alia*, the marketplace for visual content.

103. Stability AI's unauthorized use of Getty Images Marks in connection with synthetic images generated through the use of Stable Diffusion and DreamStudio constitutes trademark infringement in violation of Section 32 of the Lanham Act, 15 U.S.C § 1114(1), as such use likely has caused and will continue to cause members of the consuming public to be confused, mistaken or deceived into believing that Getty Images has granted Stability AI the right to use the Getty Images Marks and/or that Getty Images sponsored, endorsed, or is otherwise associated, affiliated, or connected with Stability AI and its synthetic images, all to the damage and detriment of Getty Images' reputation and good will.

104. Upon information and belief, Stability AI is and has been at all relevant times aware of Getty Images' prior use, and/or ownership of the Getty Images Marks. Thus, Stability AI's conduct, as described above, is willful, intentional, in bad faith, and designed specifically to

permit Stability AI to profit from such misuse in violation of Getty Images' rights in the Getty Images Marks.

105. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further acts of trademark infringement, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from infringing the Getty Images Marks.

106. Getty Images is further entitled to recover from Stability AI the damages it has sustained and will sustain as a result of the unlawful acts alleged above, together with the profits obtained by Stability AI. The amount of such damages and profits cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

107. Getty Images is entitled to recover treble damages or profits, whichever is greater, for Stability AI's use of a counterfeit mark.

108. Getty Images is entitled, at its election, as an alternative to an award of actual damages and profits earned by Stability AI, to recover statutory damages of up to $2,000,000 per counterfeit mark used.

109. Getty Images is entitled to recover its full costs and attorneys' fees in prosecuting its claims for trademark infringement.

CLAIM V

Unfair Competition in Violation of Section 43(a) of the Lanham Act, 15 U.S.C. § 1125(a)

110. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 109 above.

111. Stability AI's unauthorized use of the Getty Images Marks in the United States in connection with synthetic images generated through the use of Stable Diffusion and DreamStudio constitutes unfair competition and false designation of origin in violation of Section 43(a) of the Lanham Act, 15 U.S.C § 1125(a), as such use likely has caused and will continue to cause members of the consuming public to be confused, mistaken or deceived into believing that Getty Images has granted Stability AI the right to use the Getty Images Marks and/or that Getty Images sponsored, endorsed, or is otherwise associated, affiliated, or connected with Stability AI and its synthetic images, all to the damage and detriment of Getty Images' reputation and good will.

112. Upon information and belief, Stability AI is and has been at all relevant times aware of Getty Images' prior use, and/or ownership of the Getty Images Marks. Thus, Stability AI's conduct, as described above, is willful, intentional, in bad faith, and designed specifically to permit Stability AI to profit from such misuse in violation of Getty Images' rights in the Getty Images Marks.

113. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further acts of trademark infringement, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from competing unfairly with Getty Images.

114. Getty Images is further entitled to recover from Stability AI the damages it has sustained and will sustain as a result of the unlawful acts alleged above, together with the profits

obtained by Stability AI. The amount of such damages and profits cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

115. Getty Images is entitled to recover treble damages or profits, whichever is greater, for Stability AI's use of a counterfeit mark.

116. Getty Images is entitled, at its election, as an alternative to an award of actual damages and profits earned by Stability AI, to recover statutory damages of up to $2,000,000 per counterfeit mark used.

117. Getty Images is entitled to recover its full costs and attorneys' fees in prosecuting its claims for unfair competition and false designation of origin.

CLAIM VI

Trademark Dilution in Violation of Section 43(c) of the Lanham Act, 15 U.S.C. § 1125(c)

118. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 117 above.

119. The Getty Images Marks are distinctive and famous.

120. Stability AI has used the Getty Images Marks in commerce in the United States, and Stability AI's commercial use of the Getty Images Marks commenced after those marks became famous.

121. Stability AI's use of the Getty Images Marks on lower quality, and in some cases bizarre or grotesque images, dilutes the quality of the Getty Images Marks by blurring or tarnishment. Upon information and belief, Stability AI's use of the Getty Images Marks on lower quality, and in some cases bizarre or grotesque images, has been and continues to be knowing, willful, and in bad faith.

122. Stability AI's unauthorized use of the Getty Images Marks in connection with lower quality synthetic images generated through the use of Stable Diffusion and DreamStudio constitutes trademark dilution in violation of Section 43(c) of the Lanham Act, 15 U.S.C § 1125(c).

123. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further acts of trademark dilution, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from diluting the Getty Images Marks.

124. Getty Images is further entitled to recover from Stability AI the damages it has sustained and will sustain as a result of the unlawful acts alleged above, together with the profits obtained by Stability AI. The amount of such damages and profits cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

125. Getty Images is entitled to recover its full costs and attorneys' fees in prosecuting its claims for trademark dilution.

CLAIM VII

Deceptive Trade Practices in Violation of Delaware's Uniform Deceptive Trade Practices Act

126. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 125 above.

127. Stability AI's unauthorized use of the Getty Images Marks in connection with synthetic images generated through the use of Stable Diffusion and DreamStudio constitutes a deceptive trade practice in violation of Delaware law, as such use likely has caused and will

continue to cause members of the consuming public, including in Delaware, to be confused, mistaken or deceived into believing that Getty Images has granted Stability AI the right to use the Getty Images Marks and/or that Getty Images sponsored, endorsed, or is otherwise associated, affiliated, or connected with Stability AI and its synthetic images, all to the damage and detriment of Getty Images' reputation and good will.

128. Upon information and belief, Stability AI is and has been at all relevant times aware of Getty Images' prior use, and/or ownership of the Getty Images Marks. Thus, Stability AI's conduct, as described above, is willful, intentional, in bad faith, and designed specifically to permit Stability AI to profit from such misuse in violation of Getty Images' rights in the Getty Images Marks.

129. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further deceptive acts, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from engaging in deceptive trade practices.

130. Getty Images is further entitled to recover from Stability AI treble the damages it has sustained and will sustain as a result of Stability AI's acts in violation of Delaware law. The amount of such damages cannot be fully ascertained by Getty Images at present but will be established according to proof at trial.

131. Getty Images is entitled to recover its full costs and attorneys' fees in prosecuting its claims for deceptive trade practices.

CLAIM VIII

Trademark Dilution in Violation of Section 3313 of the Delaware Trademark Act

132. Getty Images realleges and incorporates by reference herein the allegations set forth in paragraphs 1 through 131 above.

133. The Getty Images Marks are distinctive and famous.

134. Stability AI has used the Getty Images Marks in commerce, and Stability AI's commercial use of the Getty Images Marks commenced after those marks became famous.

135. Stability AI's use of the Getty Images Marks on lower quality, and in some cases bizarre or grotesque images, dilutes the quality of the Getty Images Marks by blurring or tarnishment. Upon information and belief, Stability AI's use of Getty Images Marks on lower quality, and in some cases bizarre or grotesque images, has been and continues to be knowing, willful, and in bad faith.

136. Stability AI's unauthorized use of the Getty Images Marks in connection with lower quality synthetic images generated through the use of Stable Diffusion and DreamStudio constitutes trademark dilution in violation of Section 3313 of the Delaware Trademark Act.

137. As a direct and proximate result of Stability AI's wrongful conduct, Getty Images has been substantially and irreparably harmed in an amount not readily capable of determination and, unless permanently enjoined from further acts of trademark dilution, Stability AI will cause additional irreparable harm for which there is no adequate remedy at law. Getty Images is thus entitled to permanent injunctive relief preventing Stability AI, its agents, affiliates, employees and all persons acting in concert with it from diluting the Getty Images Marks.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff Getty Images respectfully requests judgment in its favor and against Defendants Stability AI as follows:

A. Finding that Stability AI has infringed Getty Images' copyrights;

B. Finding that Stability AI's copyright infringement was willful;

C. Finding that Stability AI has provided false copyright management information;

D. Finding that Stability AI has removed or altered copyright management information;

E. Finding that Stability AI has infringed Getty Images' trademarks;

F. Finding that Stability AI has diluted Getty Images' trademarks;

G. Finding that Stability AI has tarnished Getty Images' trademarks;

H. Finding that Stability AI's trademark infringement, unfair competition, trademark dilution, and deceptive trade practices were willful and in bad faith;

I. Finding that there is a substantial likelihood that Stability AI will continue to infringe Getty Images copyrights and trademarks unless enjoined from doing so;

J. Issuing a permanent injunction enjoining Stability AI and its agents, servants, employees, successors and assigns, and all persons, firms and corporations acting in concert with it, from directly or indirectly infringing Getty Images' copyrights, from providing false copyright management information, from removing or altering Getty Images' copyright management information, and from infringing, diluting, or tarnishing Getty Images' trademarks;

K. Ordering the destruction of all versions of Stable Diffusion trained using Getty Images' content without permission;

L. Ordering Stability AI to provide a full and complete accounting to Getty Images for Stability AI's profits, gains, advantages, and the value of the business opportunities received from its infringing acts;

M. Entering judgment for Getty Images against Stability AI for all damages suffered by Getty Images and for any profits to or gain by Stability AI attributable to its infringement of Getty Images' copyrights and its acts in violation of 17 U.S.C. § 1202

N. Entering judgment for Getty Images against Stability AI for all damages suffered by Getty Images for any profits to or gain by Stability AI attributable to its infringement and dilution of Getty Images trademark and its unfair competition and deceptive trade practices in amounts to be determined at trial, with the greater of such damages and profits trebled;

O. Entering judgment for Getty Images for statutory damages for Stability AI's willful acts of copyright infringement, its provision of false copyright management information, and its removal or alteration of Getty Images' copyright management information;

P. Entering judgment for Getty Images for statutory damages for Stability AI's willful acts of trademark infringement and unfair competition;

Q. Awarding Getty Images its costs and reasonable attorneys' fees;

R. Awarding Getty Images pre-judgment and post-judgment interest to the fullest extent available; and

S. Granting such other and further relief as the Court deems just and proper.

DEMAND FOR JURY TRIAL

Plaintiff Getty Images demands a trial by jury on all issues so triable.

Dated: March 29, 2023

YOUNG CONAWAY STARGATT & TAYLOR, LLP

/s/ *Tammy L. Mercer*
Tammy L. Mercer (No. 4957)
Robert M. Vrana (No. 5666)
1000 North King Street
Wilmington, Delaware 19801
(302) 571-6600
tmercer@ycst.com
rvrana@ycst.com

WEIL, GOTSHAL & MANGES LLP
Benjamin E. Marks (admitted *pro hac vice*)
Jared R. Friedmann (admitted *pro hac vice*)
Melissa Rutman (admitted *pro hac vice*)
767 Fifth Avenue
New York, New York 10153
(212) 310-8000
benjamin.marks@weil.com
jared.friedmann@weil.com
melissa.rutman@weil.com

Attorneys for Plaintiff Getty Images (US), Inc.