I previously submitted an incomplete response due to time constraints, but have now added my answers to the other questions that I would like to answer. This document is my complete comment to the US Copyright Office on issues related to AI and copyright. Question numbers highlighted in yellow are newly added or contain revisions compared to the previous version.

Note: I am a visual artist that prefers to stay anonymous to protect my employment status. Most of my comments relate to image-generating AI such as Stable Diffusion, although for question #7.1, my generic statements about the basic idea behind AI training and how weights are updated (gradient descent) also apply to other forms of generative AI.

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?

To me, generative AI is an advanced form of data approximation and interpolation. It can be viewed as a way of analyzing the underlying trends and features of the training dataset, and thus has many uses when this dataset contains factual information that humans wish to make sense of. Examples include molecular design, protein folding, traffic prediction and epidemic spread prediction.

However, when extended to the realm of human expressions, data interpolation becomes a mimicry of human culture and devalues individual expressions through generalization. I work in the visual arts industry. When generative AI gained popularity in 2021, I was initially excited and thought it had the potential to easily generate filler content in large projects, which would allow artists to focus on the key areas that deserved more attention. However, as time progresses, not only are the benefits not realized, but this technology has done way more harm to the creative industries than good. It is now working like a parasite on human creatives, draining the hosts, displacing jobs, obscuring authorship and diluting authentic expressions. I will elaborate on this in my answer to question #2. This is especially problematic when the training dataset contains copyrighted materials, which I will explain in my answer to question #7.1.

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?

Yes. I work as a senior concept artist in the gaming industry during the day and as an independent artist/illustrator at night, so my answer will be divided into different sections:

I. Within game studios:

Companies including the one that I work in are starting to train AI models based on our (work-for-hire) artwork with the intention of specifically mimicking our art styles. This is done for the potential of quicker and/or mass production of artwork that fits into our project aesthetics, but it is deeply concerning to me in many ways.

1. Mimicry of artistic identity

Stylistic mimicry can lead to problems for artists on a personal level just like deepfake does to people in real life. When a human artist produces an artwork, even when we try to adjust our styles to fit the needs of a project, traces of our personal habits and stylistic signatures will remain and the people familiar with them can often tell who painted which image. In this sense, our art styles are linked to our identities in people's perceptions, and having an AI model of that, which the company has full control over, is both risky and disturbing to us as individuals. We signed up for our jobs before the prevalence of generative AI to produce individual art pieces, not to sell our artistic identities to the companies. However, we currently have no legal grounds to object to this.

2. Job displacement at a cost to the public

These AI models have the potential to permanently displace a large portion of artists from our job positions. While some may argue that job displacement as a result of technological innovation is a natural process, the problem this time is that it will also downgrade the quality of entertainment the public receives. Even though AI generations cannot reach the bespoke quality of human creations, the ease and low cost of using AI still makes it a desirable option for businesses from a profitability standpoint. More importantly, in the long term, I foresee a decrease of new artists entering this field due to bad career prospects. This is problematic because while generative AI is good at producing "good enough" quality images at a large scale, it does not have the ingredients to push boundaries, as it is fitted on existing content. (Just to clarify, if done well, it is possible for generative AI to produce more intricate details and finer renderings, or even to come up with new aesthetics through combining existing ones, but it does not go beyond the status quo in the "interpretive" aspect of artistic expressions. For instance, an AI model trained on realism will not be able to generate impressionistic artwork, as this leap of visual interpretation requires the mind of a human.) The entertainment space will likely be stagnantly filled with profitable but generic content unless we encourage new human artists with their own voices and interpretations to come into the scene.

II. Among indie artists and in the general creative ecosystem:

Many studio artists like myself also practice art after hours, either posting personal work online or selling them in marketplaces. This is the opportunity for us to project our creative voices to the world, market ourselves and inspire each other. It is also through this avenue that I get to experience the lives of indie artists and freelancers, who are also facing problems caused by generative AI.

1. Parasitic behavior and performance punishment:

Within the last year, I have had multiple occasions where my artwork was mistaken to be AI generated both online and in person, possibly due to the aesthetics resembling certain classes of AI images at first glance. This is disheartening to me as I put in tremendous effort telling a human story through each image. One may suggest I look for alternative art styles, but the problem is, it takes years for an artist to develop a new style, but fine tuning an AI model to fit that style can take as little as 10 images. In this sense, the technology works like a parasite. The further irony is that the more appealing an artist's body of work is, the more likely these works will be used as training materials. Artists like myself do not know what to look forward to because putting in more effort honing our craft only means a higher chance of getting our personal voice appropriated. This kind of performance punishment is detrimental to the creative community.

2. Devaluation of existing art styles through overexposure:

Before generative AI, this was a negligible issue that may have even played a positive role in transforming the creative landscape over time, but it is now massively amplified by generative AI to a point that it is disruptive and unsustainable. In general, for an artwork to appeal to the public, it needs to have some level of familiarity to the audience to form a connection, and an element of novelty that piques interest. Without this balance, the work becomes either too generic or too difficult to grasp, thus limiting its commercial success. This applies to both subject matters and aesthetics. The problem with generative AI is that it latches onto a successful art style and massively produces similar images that flood the content space, quickly overwhelming the audience with this aesthetic and devaluing it. Based on my experience, many people turn away from an artist's work if they recognize a similar style in AI images. This significantly harms the artists whose works are heavily influential in AI training, even if the AI generations are neither directly plagiarizing (by current standards) nor used for commercial gains by the prompters.

3. New artists drowned and established artists start to paywall:

Due to its high output, the amount of AI generation in the content space will likely exceed human creations in the future. Finding the seed to exposure for a new artist was already a difficult task before, and it will only get worse onwards. On the other side of the spectrum, artists with a large enough follower base are incentivized to paywall artwork that would otherwise be free for public viewing, out of fear of having them scraped into training databases. Taken together, not only do these factors obstruct the flow of quality information that the creative community relies upon for inspiration, but also weaken the human connection that arts and culture are meant to bring to the general public.

III. As a consumer:

Most artists are also consumers of creative content. Personally, I also interact directly with consumers in marketplaces, and because of this, I can speak from their perspective on issues related to generative AI.

1. "Bad money drives out good"

As mentioned previously, my artwork tends to be quickly dismissed by those who mistaken them to be AI generated. This is because many consumers perceive AI generated content to hold less value than human creations, due to their lack of (perceived) association with skill, effort and human connections. However, generative AI is now capable of producing images that are difficult for the layperson to tell without careful examination. One recent example is when the voice actress Grey DeLisle brought an art print of her voiced character, Daphne, only to find out on social media that it was AI generated and demanded refunds. This is a problem that is increasingly plaguing art marketplaces. Without a way for people to pick out authentic products from lookalikes, consumers may lose trust in the current forms of creative content overall.

2. Loss of authorship information and origin stories

When an artwork piques my interest, I would like to know the author and story behind it, what the sources of inspirations are and how artistic choices are made. Not only does this pay respect to the author, looking into the experience of another person is a fundamental part of human learning and helps us form empathic relationships with the world. However, AI generations do not contain this information. Due to their statistical fitting nature, rather than showing how individual humans express themselves, AI outputs are more akin to a general statement of "this is how others express themselves", yet no information is given toward who the "others" are nor why they make such expressions.

I would like to illustrate this point with a simplified scenario. Say, for example, if two artists were asked to paint "a neighborhood on a summer day", one of them might relate to sensory experiences such as heat, bright sunlight, and the scent of summer vegetation, hence painting a warm, saturated and overexposed street scene. Another artist might relate back to her childhood during summer vacations and paint a cheerful imaginary neighborhood playground. Both of these are individual expressions that embody each artist's personality, habits, life story and world interpretation. Now, if a program takes the two images as data points and creates an "average" of the two, the resultant image, despite still looking like an artist's creation following the same prompt, obscures the original intentions yet provides very little added value. Of course, this is an oversimplification of how generative AI works, but the objective is the same: to create a fitted model approximating existing data, and generate new data based on interpolation. In the simplified scenario, one can still trace back to the original artists, but when the dataset extends to all the content available on the internet, tracing becomes impossible, yet for any set of existing human data, there can be an infinite number of "generative" interpolations that confuse the audience. As a consumer, I do not look forward to exploring a content space like this.

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.

AI Art and Its Impact on Artists
https://dl.acm.org/doi/10.1145/3600211.3604681

4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries that relate to copyright and AI that should be considered or avoided in the United States? (40) How important a factor is international consistency in this area across borders?

My understanding is that under EU's AI Act, companies training generative AI have to publish summaries of copyrighted data used for training and disclose content that was generated by AI. As a consumer, I like this direction toward transparency and hope more countries can follow. At the same time, I believe this is still not enough to protect creative communities and authentic human expressions, and that regulations need to be in place to control what can be used as training data.

I also think international consistency is important because data is easily accessible across borders. I believe this is why the Berne Convention existed for copyright. Otherwise, AI companies can simply move to countries with less regulations and offer services back to countries with more regulations. However, this is not to say that it is pointless to regulate unless other countries are doing the same, as this will simply become a race to the bottom and create an undesirable outcome for everyone. Rather, I hope the US can leverage its international influence and lead the world in building a safe environment for human intellectual activities in the age of AI.

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.

Yes. For visual artists, their art style is an embodiment of their personality, taste, and habits of expression that are individualistic and signatory. It is often perceived to be the artist's identity in the creative space. However, unlike an actor's physical image or a singer's voice, a visual artist's style is protected neither as part of the artist's personality rights nor under copyright law, and this often leads to abuse by users of generative AI.

Before the rise of AI, stylistic mimicry was not a major issue because human artists have very little incentive to mimic the styles of others, as the time and effort needed for successful mimicry can better be spent finding and honing the artist's own creative voice. However, generative AI makes it possible for anyone to use the style of another artist with ease. This is problematic for the original artist both economically—the creative voice that took them years to develop is being devalued into a "stock"—and personally—their identity in the artistic space can now be associated with projects that they didn't approve of.

Some are suggesting making art styles copyrightable. This would cause many problems practically. For example, in collaborative projects such as game development, multiple artists have to match styles to a certain extent for consistency. (Note that, as opposed to the mimicry scenario I mentioned above, this kind of style matching is done with permission and is rarely precise.) Another problem with copyrighting art styles is that large corporations can find ways to achieve style monopolies through hiring human artists and asking them to assign rights to the company. This is not good for the creative economy overall and will destroy the livelihoods of artist employees when they leave the company.

I believe art styles should be protected in a separate category that specifically addresses the problems introduced by generative AI. I do not work in law, so I can only speak of a human's natural instincts. Intuitively speaking, there are different "degrees" of personal identifiers. Conventional ones, such as a person's name and physical likeness, are directly linked to the individual's identity and should be protected by publicity rights. At the same time, there is another category of identifiers that, while not rigorously tied to one's physical identity, "suggests" authorship in public perception and can still harm the individual when mimicked and misused—a problem that is now exacerbated by machine learning. This includes a person's handwriting, an artist's style, and other personal habits of creative expression. This category of characteristics is a mixture of personal identity and intellectual accomplishment, and should be protected in a special way. Because art styles can be ambiguous (two artists can naturally draw similarly to each other) and changing over time (an artist's style may evolve), one cannot simply assign a certain "look" to an artist, but rather, a better way of protecting art styles is to forbid certain actions from being performed on them. Specifically, these are actions introduced in the age of AI that are style-extracting and style-devaluing in nature, including but not limited to 1) directly using an artist's work for machine learning without permission (even if it is mixed with other images), 2) having a human mimicking the style of another artist for the purpose of machine learning, and 3) using an AI model trained on an artist's style for a purpose or duration of usage different than what was agreed upon. None of these affect human artists learning from and influencing each other, as I believe any new legislation should not have to majorly alter the way creatives behave when AI is left out of the equation.

Regarding getting permission from copyright holders for AI training, I believe it is equally, if not more, important to get consent from the "physical" authors of the artwork who may have signed away their rights to that particular artwork. In the broader sense, style extraction can do even more harm to an artist than plagiarism or distortion of individual copyrighted pieces, and should be addressed seriously in order to have a sustainable creative ecosystem. I believe there should be some unwaivable form of truth connecting an artwork to the physical author whose human imprint is contained within the work, so that creatives can object to having their styles extracted through machine learning regardless of whether or not they are the copyright holders (or whether or not they waived the moral rights to their creations).

Another possible scenario is that in an employment relationship, employers will try to have employees sign away all their rights and agree to have their work used for AI training. I believe blanket terms asking for generic and perpetual usage of an artist's style in an employment contract should be outlawed. This is due to its potential to cause irreversible damage to the artist's livelihood, even after the employment relationship ends, through a single step of negligence or in a scenario of power imbalance in a bad economy. Rather, permission should only be obtained with clear statements of the purpose and the duration of usage of the AI model to be trained, and that the model should be destroyed once the term ends.

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)?

Stable Diffusion was trained on subsets of data within LAION-5B, an uncurated dataset that was constructed through web scraping:

1. Common Crawl, a non-profit organization, built a web archive by crawling the web since 2008 and publishing the results every month.

2. LAION used this archive to compile image-text pairs, by extracting (from HTML) image alt-texts and corresponding image URLs, and temporarily downloading the images for analysis.
3. LAION then used automated methods to filter out low quality data, resulting in 5.85 billion image-text pairs that formed the LAION-5B dataset.
4. The dataset can be classified by language, resolution, and other metrics. One particular metric of interest is the "aesthetic rating", which is a score given by LAION's own AI model that predicts how much people "like" an image on a scale of 1-10. Unfortunately, "copyright" or "image licensing" are not within the list of classifications.

(Sources: https://arxiv.org/pdf/2210.08402.pdf, https://laion.ai/blog/laion-aesthetics/)

Specifically, Stable Diffusion used these subsets within LAION-5B:

- LAION-2B-EN: 2.3 billion English-captioned images
- LAION-High-Resolution: 170 million images greater than 1024x1024 resolution
- LAION-Aesthetics-v2-5+: 600 million images within LAION-2B-EN that have an aesthetic score of 5 or higher

Note that while LAION itself is a non-profit organization, it received funding from Stability AI, the company behind Stable Diffusion.

It is unclear what datasets Midjourney and DALL-E were trained on, but based on the aesthetics of their output it is very likely that they used something similar to those by Stable Diffusion. (Evidence in the Lawsuit Andersen *et al. v* Stability AI Ltd. *et al.* seems to suggest that Midjourney was trained on LAION-400M, which was also scraped from the Internet: https://laion.ai/blog/laion-400-open-dataset/)

Adobe's generative features were trained using Adobe stock images. Other stock image services such as Shutterstock and Getty Images seem to be doing something similar. However, it is unclear how much the stock image contributors are getting compensated, or whether or not they consented to having their stock images used in AI training.

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?

Because LAION-5B was built from web scraping, copyright owners have no knowledge of whether or not their work was used unless they find them within the dataset on their own. Personally, using the site haveibeentrained.com, I could find all the artwork I posted in my ArtStation portfolio before mid-2018, but I did not give anyone permission to use those artwork for AI training.

Andy Baio and Simon Willison conducted a domain sampling experiment using data from 12 million images (2% of the high-aesthetic dataset) used to train Stable Diffusion, and found over a million images coming from Pinterest (which contains many unauthorized reposts of artists' works), 698k images from Fine Art America, 244k from Shopify, 232k from Smugmug, 90k from Redbubble, 67k from DeviantArt and 47k from Etsy. Most of these images, especially those from shopping sites and portfolio sites, should have copyright protection in some ways. More information on their experiment can be found in this blogpost: https://waxy.org/2022/08/exploring-12-million-of-the-images-used-to-train-stable-diffusions-image-generator/

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.

To answer this question, I will first provide an overview of how generative AI works and how weights are updated during training, and then use Stable Diffusion as an example to show how training materials are used in detail. I will bring up my views on how these activities affect the rights of copyright owners at relevant places.

The Overall Picture:

The basic idea behind generative AI is to approximate the data distribution of the training materials, and then sample from this to get "new" data points that appear to belong to the same distribution. To form this approximation, parameters (weights) affecting how an output is produced from a neural network are updated iteratively to minimize a loss function—a mathematical expression that measures some form of a "difference" between the current approximation and the training data. (Each generative model expresses the loss function differently following its training objective. For example, a variational autoencoder, which trains by compressing images into abstract representations and then reconstructing the images, writes the loss function as the reconstruction error plus a term to regularize the abstract space. A diffusion model, which attempts to form an image from noise by stepwisely denoising it, writes the loss function as the difference between the "predicted noise" to be removed and the real noise applied on each training image. Other AI frameworks use other objectives, but the underlying idea is the same: the training dataset forms the ground truth that the AI model tries to minimize its difference from.) The reason why sampling from this creates "new data" rather than regurgitates training data is that design tricks, such as treating data points as probability distributions, are used to make the approximation more continuous. This way, sampling becomes some form of data interpolation, yielding results that look like they can fit into the overall training distribution rather than forcing them to become exact copies of training data points.

This method of approximation and sampling is mimicry in nature, because its very objective is to make its output blend into the training dataset, or in other words, to mimic the “look and feel” of the training materials. A common analogy of this often used by AI companies to justify their products is “a student learning from a teacher”. This is extremely misleading as no human student learns solely from the creation of others, let alone minimizing statistical differences between their work and the teacher’s body of work. Rather, human artists incorporate their own personality, habits, experience and world interpretation in the process, and aim to find unique artistic expressions different from what is already existing. In terms of copyright implications, AI generations are derivative works of copyright-protected materials that compete with them in the same market. Even if not sold commercially, they devalue the training materials by exposing the audience to large quantities of similar-looking content. My answer to question #2 contains an elaboration on this.

Gradient Descent:

During training, weights are updated by a process called gradient descent, which works as follows:

1. Start from a random set of weights
2. Calculate how the loss function changes with respect to each weight (gradient)
3. Adjust each weight by a small increment in the direction that makes the loss function decrease the most
4. Repeat steps 2 and 3 until the loss reaches a minimum

In practice, doing this for an entire training dataset is computationally expensive, so a modified version of it, called stochastic gradient descent, is used. Here, instead of computing the loss gradient over the entire dataset, one randomly selected example (or a small batch) of training data is used in each iteration. The loss function can still be minimized this way with more iterations. While what I am about to describe is not meant to be a realistic visual representation of the training process, but in a conceptual way, this is analogous to starting from a poor image and comparing that to an existing artwork, nudging every detail of the poor image slightly toward the artwork, then switching to a different training artwork, nudging again, and repeating enough times until the result looks like an artwork itself. Meanwhile, record down the parameters that allowed this process to happen and call that “learning”. To me, not only is this a subtle act of "copying" (just in small increments and mixed over millions of examples), but more importantly, no human learns this way. Because of this, I believe machine algorithms should not have the same “learning rights”, just like how cameras do not have the same observation rights, as humans do.

Stable Diffusion:

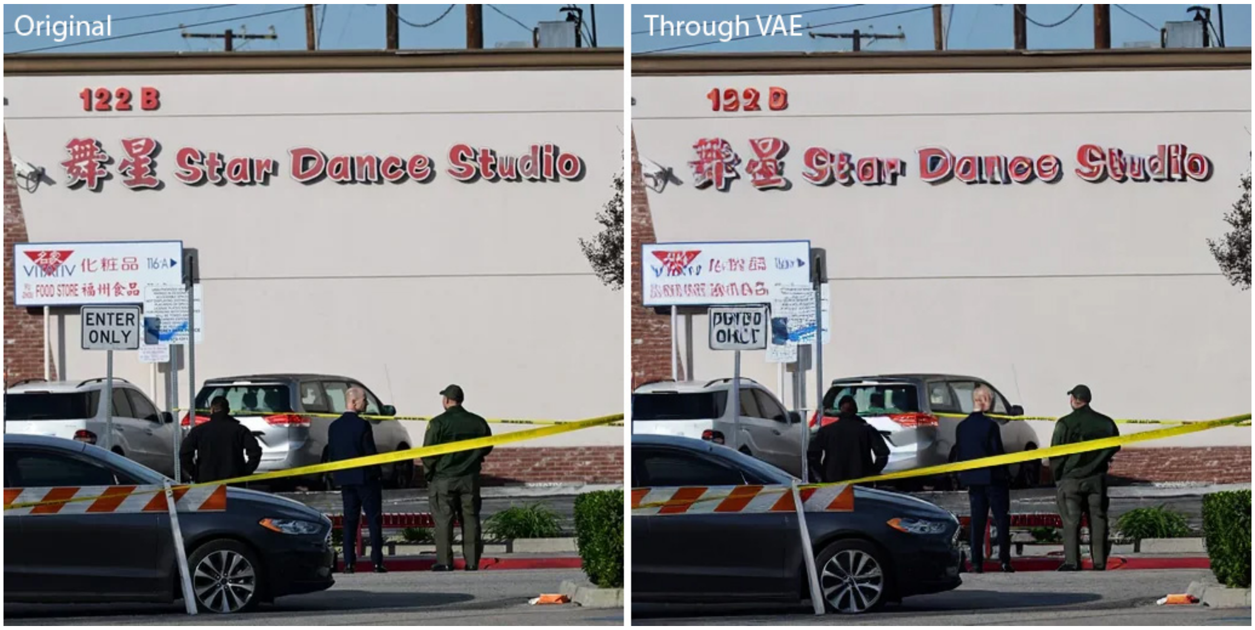
I would like to use Stable Diffusion as an example to illustrate how training materials are used in more detail. Before I start, I would like to point out that training uses image data rather than images in their visual form, and during each training loop the images do not have to be visually represented for a human to view. However, it should be fair to say that using image data for computation is the same as using the images themselves. Likewise, mathematical operations should be treated as physical actions based on their underlying meaning.

Stable Diffusion contains three parts: a variational autoencoder, a latent diffusion model, and a conditioning component. (Paper: https://arxiv.org/pdf/2112.10752.pdf)

Variational Autoencoder (VAE):

The purpose of the variational autoencoder (VAE) is to compress starting images into abstract representations of them in a “latent space”, and later on to reconstruct images from this space. The reason for doing this is that the latent space removes unimportant pixel-level details and groups similar contents together. (For example, all information related to “eyes” will be positioned close to each other in this space.) This makes the starting materials smaller and easier to work with for the subsequent step. Again, not intended to be a rigorously truthful visualization, but a rough analogy is to break each starting image into small pieces, put pieces that are similar to each other close together and give them a label (such as “eye”), so that we can work with the labels rather than the image pieces in the next step, while at the same time have a decoder that can turn the labels back into image pieces in the end.

The VAE thus has two components: an encoder that compresses the image, and a decoder that reconstructs the image, both of which are trained together. The training process involves feeding the training images into the encoder and letting the decoder make reconstructions, which are then compared back to the starting images. Weights within the VAE neural network are iteratively updated to minimize a loss function that includes a pixel-level reconstruction error between the output and the starting image. (To clarify, this loss function also contains other terms, but they do not affect the main point of this section.) Since training starts off with random weights, the decoder output is initially meaningless, but as the model iteratively updates its weights, the output will gradually start to resemble the training materials. To get a sense of the reconstruction quality, below are images posted by a reddit user (Wiskkey), where an image was passed through the VAE of a fully trained Stable Diffusion model (please note that this is a new image, not a training image, but it illustrates my point about the training end result):



(Source: https://www.reddit.com/r/StableDiffusion/comments/10lamdr/stable_diffusion_works_with_images_in_a_format/)

The final training iterations are clearly acts of making copies of training materials. Even earlier iterations, where weights are not fully adjusted, can still be viewed as failed copying attempts (possibly at the same time violating the moral rights of the image authors). Some may argue that the training process is done behind the scenes, and its purpose is to "learn" how to reconstruct images rather than to make use of the reconstructed images themselves. However, this *act* of copying is an irreplaceable step to creating a usable AI model, and the copyrighted materials involved also cannot be replaced by low quality data without affecting the quality of the model's output. Therefore, the commercial usability of Stable Diffusion (and other similar models) is dependent on the action of copying copyright-protected materials.

Some may also argue that the VAE merely serves to translate images into and out of their latent representations, while the diffusion model is where the "magic" happens. I will address this in the next section.

Latent Diffusion Model:

The purpose of the latent diffusion model is to introduce more variability in the generated content. It is performed in the latent space to save computation costs and to allow for conditioning. However, the previous section has shown that the decoder can faithfully reconstruct images from the latent space, so I will describe the diffusion process as if it was operated on the images directly for simplicity.

The way a diffusion model generates images is by stepwisely forming an image from pure noise. This is a reversal of the forward "diffusion" process, where noise is added to an image, one small amount at a time, until the image becomes unrecognizable. To achieve this reversal, the model only needs to predict the "noise" that was added in each step, and subtract that from the noisy image to get a cleaner result. The diffusion model thus needs to be trained for its ability to come up with the "right noise" to be subtracted.

The training of a diffusion model follows these steps:

1. Randomly choose an image from the training dataset.
2. Randomly choose a timestep that corresponds to a noise level. (This is because the denoising process happens in a stepwise manner.)
3. Randomly generate a noise sample. (Two images made of pure noise may look indistinguishable to humans, but the individual bits of noise may still be arranged differently, making each noise sample unique.)
4. Apply this noise to the training image at the correct level chosen from step 2, and using this "noisified" version of the training image, let the neural network make a prediction on the noise sample that was added.
5. Find the difference between this and the actual noise sample from step 3. This is the loss function that needs to be minimized.
6. Update weights and repeat from step 1, until the loss is minimized.

(Algorithm 1 from Ho *et al.* https://arxiv.org/pdf/2006.11239.pdf)

On the surface level, this may seem like a process of predicting "noise" rather than training images. However, given that the "noisified" training image that the neural network uses to make its prediction is the sum of a noise sample and a training image, predicting the noise is mathematically equivalent to predicting the training image. Therefore, each training iteration is essentially an attempt at reconstructing a training image from an arbitrarily noisified version of it, with an objective of minimizing the difference between the reconstruction and the original. Therefore, just like that of the VAE, training the diffusion model involves the repeated action of copying training materials, from initial failed attempts to later successful attempts. As for the scale at which this copying was performed, here are the numbers of training steps (copying attempts) Stable Diffusion V1 took:

sd-v1-1: 237,000 steps on LAION-2B-EN; 194,000 steps on LAION-High-Resolution
sd-v1-2: Resumed from sd-v1-1; 515,000 steps on LAION-Aesthetics-v2-5+
sd-v1-3: Resumed from sd-v1-2; 195,000 steps on LAION-Aesthetics-v2-5+
sd-v1-4: Resumed from sd-v1-2; 225,000 steps on LAION-Aesthetics-v2-5+
sd-v1-5: Resumed from sd-v1-2; 595,000 steps on LAION-Aesthetics-v2-5+
sd-inpainting: Resumed from sd-v1-5; 440,000 steps on LAION-Aesthetics-v2-5+

(Source: https://huggingface.co/runwayml/stable-diffusion-v1-5)

Because these models were trained at a batch size of 2048, the number of steps should be multiplied by this number to get the actual quantity of images being copied. If we consider the actions alone rather than their success or failure, this is copyright infringement at an unprecedented scale.

Conditioning Component:

The purpose of conditioning is to allow for image generation that follows guidance provided by the users in the forms of text, images, or other modalities. This is achieved by exposing the latent diffusion model to conditioning signals during training. In the case of text-to-image generation, this conditioning signal is in the form of text embeddings, which are vectors that contain information on what words are closely related to each other and how they are associated with imagery. This information on text relationship comes from a separately trained AI model: Stable Diffusion V1 uses OpenAI's CLIP text encoder, which was "trained on a dataset of 400 million (image, text) pairs collected from a variety of publicly available sources on the Internet" (https://arxiv.org/pdf/2103.00020.pdf). Stable Diffusion V2 uses OpenCLIP, which was trained by LAION on LAION-2B (https://laion.ai/blog/large-openclip/). Clearly, these models were not trained with permission from copyright owners either.

The conditioning signal is incorporated into the diffusion model through a cross-attention mechanism. The setup has similarities to that of ChatGPT and is beyond the scope of this section, but due to this design, the text portion of Stable Diffusion also contains visual information, and fine-tuning by the end user often happens by modifying these cross-attention layers.

End User Fine-Tuning:

It is easy for an end user to customize Stable Diffusion through fine-tuning. A popular method to do so is through LoRA (Low-Rank Adaptation), which uses the fact that weight matrices (number arrays) can be mathematically broken down into smaller matrices that are easier to adjust. The user can then use customized datasets to create small weight matrices to be inserted into the pre-trained Stable Diffusion (often in the text encoder) to add information, such as specific art styles, personal likeness, and conceptual ideas. From an operational perspective, training a LoRA involves gathering ten to a few hundred relevant images, creating a caption for each, setting learning parameters, and running an existing training script. Afterwards, the LoRA can be used by the user directly, uploaded onto platforms such as Civitai for others to download, or to be incorporated directly into other online AI services. While training a LoRA may take some curatorial efforts and trial and error setting learning parameters, most of what contributes to the artistic qualities of the customization comes from the training images. However, as LoRA training can easily be performed by anyone, tracking down what training materials were used would be difficult unless distribution platforms such as Civitai require their disclosure.

There also exists other methods of fine-tuning, such as by using DreamBooth or textual inversion, but they all pose the same problems.

Summary:

My answer to this question was long, so I will provide a summary of my main points:

- Generative AI works through statistical fitting, which not only defies the very idea of creativity, but also devalues the training materials through competition and over-exposure.
- During training, weights are updated by gradient descent, a process that, to me, is a subtle form of copying and is also very different from human learning. Because of this, I believe machines should not have the same rights to “learn” from human output as other humans do.
- The training process of image-generating AI, such as Stable Diffusion, involves millions of attempts at reconstructing the training images, from initial failed attempts to later successful ones. Considering actions rather than results, this is copyright infringement at a massive scale.
- End users can easily fine-tune pre-trained models using small customized datasets. It is very difficult to track what went into these unless some kind of regulation is in place.

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

As described in my answer to question 7.1, during training, the trainable parameters (weights) within the neural network are updated according to the training objective (loss function), and with these updated weights, the models are able to generate the desired kinds of output using their sampling methods. Basically, what is gained from the training process are large sets of numbers (hundreds of millions for Stable Diffusion V1) that are meaningless to humans and only make sense operationally when plugged into the neural network that they were trained in. For Stable Diffusion, these weights are packaged in checkpoint files that can be publicly downloaded. AI services such as Midjourney keep their weights secret.

(Optional read) What happens during sampling (inferencing):

To show how training and sampling relate to each other, here is a comparison of the diffusion model’s training algorithm and one of its sampling algorithms:

Training algorithm (from my answer to 7.1):

1. Randomly choose an image from the training dataset.
2. Randomly choose a timestep that corresponds to a noise level.

3. Randomly generate a noise sample.
4. Apply this noise to the training image at the correct level chosen from step 2, and using this “noisified” version of the training image, let the neural network make a prediction on the noise sample that was added.
5. Find the difference between this and the actual noise sample from step 3. This is the loss function that needs to be minimized.
6. Update weights and repeat from step 1, until the loss is minimized.

Sampling algorithm (no guidance):

1. Randomly generate a noise sample.
2. Start from the timestep that corresponds to the highest noise level.
3. Using this noise sample and timestep, let the trained neural network generate the noise that needs to be subtracted to arrive at an image. (The neural network was specifically trained to do this according to step 4 of the training algorithm)
4. Make this subtraction, but only to a degree to arrive at a noisy image (with the noise level corresponding to the next timestep). This noisy image will be used in the next round in place of the pure noise sample.
5. Repeat from step 2, but go to the next timestep, until we reach the end.
6. The final image will be clean, and that is the AI output.

(Algorithms 1 and 2 from Ho *et al.* https://arxiv.org/pdf/2006.11239.pdf)

In short, the set of weights adjusted from training allows the neural network to do its job during sampling, because the model was designed this way. The weights will be meaningless if placed out of context.

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?

I did not answer questions 8.1-8.3 due to my lack of legal knowledge, but for this question, pre-training of an image-generating AI such as Stable Diffusion uses millions to billions of images. However, the volume of material used to train an AI model should not affect the fair use analysis because not all training images affect each output equally. For example, if a prompter wants to use Stable Diffusion to create a “tree in the style of Studio Ghibli”, the resultant image will be more heavily influenced by images of trees and artwork by Studio Ghibli within the training data, and other images such as a photo of Mars will be less relevant. In other words, each output specifically affects the copyright holders of the training materials most relevant to the prompts, and considering all the active users of AI generators and different prompts that can be used, all copyright holders will likely be affected in one way or another by specific uses (but to different degrees depending on prompt popularity). Just because the overall training dataset is large doesn't dilute down the problem.

Another way to look at this is that even though a large volume of material is used for training, the AI output is also limitless in numbers and encompasses everything that was used in training.

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 owners should have to affirmatively consent (opt-in) to the use of their works for training:

1. The sheer number of AI models being trained (customized) at any given time makes it impossible for copyright owners to track where their work is being used. One can confirm this by examining the website civitai.com.

2. Most of the time, copyright owners do not know that their work is being trained on until the model is released. By then it is already too late because AI models cannot "unlearn".

3. Some copyright owners may not be alive.

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?

For fine-tuning, which is usually targeted at specific sources and only requires from ten to a few hundred images, it is definitely feasible.

For pre-training, this may be difficult due to the volume of training materials needed, but I believe it should be up to the company to figure out a way to do so if they wish to engage in such businesses to start with. It is not a good argument to say "I'm going to harm someone for my own gains because it's too difficult otherwise."

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?

The creator should have a right to object. Please see my answer to question #5.

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?

Yes. Both developers of AI models and creators of training datasets should be required to collect, retain and disclose records regarding training materials.

15.1. What level of specificity should be required?

The level of specificity should be enough for individual copyright holders to identify whether or not their work is included in the training dataset.

15.2. To whom should disclosures be made?

Disclosure should be made to the public because large datasets can potentially contain anyone's work. It should also be presented in a navigable way for consumers to view, as it holds a similar purpose as a product label or movie credits.

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?

Individual creators do not have the resources to always be kept up to date about which dataset contains their work, and there are very limited actions they can take if they do spot their work in an existing dataset. Therefore, a recordkeeping system alone is insufficient, and will have to work in conjunction with another system that obtains permission from the copyright holders before training takes place.

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

Developers of AI models should be obliged to seek permission from copyright owners before using their work as training material (thereby notifying copyright owners).

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

Legal protection for AI-generated material is not desirable as a policy matter.

At the current rate, AI generated content will outcompete human creations through sheer quantity and does not need further encouragement. In the gaming industry, many studios are already making use of AI generated images in place of human art, while being fully aware that AI generations are not copyrightable. This is because these images are so cheap to make that they can be treated as disposable. Also in practice, some are trying to circumvent the lack of copyrightability of AI-generated images by having a human painting over them and making alterations, so that it becomes undesirable for others to make use of these images directly. This is already causing job losses and unfair competition with other human creators, and having AI generated content being copyrightable will be a nightmare for human creatives.

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?

Ideally, yes. Access to truthful information is the foundation on which a functional democratic society operates. It is also important to foster consumer trust in the creative industry. However, the system will need to be quite complex to serve any practical purpose, so I'm not sure about its feasibility.

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

A truthful and thorough labeling system might only be achievable through a combined effort of the provider of the AI service, users that generate and/or distribute AI output, platforms hosting creative works, an auditing system and community tagging. I believe the AI service provider, the user and the platform should bear most of the responsibilities.

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

The boundary between human creation and AI-assisted content production is getting increasingly blurry, especially as popular software such as Adobe Photoshop incorporates generative features. If an image only contains 10% AI-generated content because the user chose to extend the canvas of a photo for a better composition, would this image be considered an AI generation? What about an artist that started with an AI-generated image and painted on top of it, eventually covering the entirety of the canvas so that the only AI-relevant aspect is the starting composition? (In fact, there are now traditional artists that use AI-generated imagery as a reference to copy from.) What about a character design sheet for a game or animation that is entirely drawn by a human, but the character *design* itself is taken from an AI generation? Any labeling or identification will need to specifically identify which parts of the image are AI-generated, or which stage of the production process was AI involved, rather than using simple "AI" or "no AI" tags.

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?

Yes. As mentioned in my answer to question #5, a visual artist's style is perceived to be the artist's identity in the creative space, and should be protected from AI mimicry. I believe any human should be eligible for such protection, just like any human's artistic creation is eligible for copyright protection. This should also hold regardless of whether or not the individual owns the copyright to their own work, as art style is part of the artist rather than part of the product. Because of this, I believe there should be some form of truth connecting an artwork to the physical author whose human imprint is contained in the work, which is separate from the transferable copyright of the particular art piece. Using human artwork for AI training should require both the permission from the copyright holder and the permission from the physical author. Failure to do the latter would be an infringement on some form of "right" that is not currently defined, but should be in place. This is also consistent with my belief that machine learning through gradient descent should not have the same access to "learning" materials as humans do.

Technical note: As long as the works of the human creator are in the training dataset, the AI system has the potential of generating in this style even if the artist name is blocked in the prompt list. Thus, I believe protection against stylistic mimicry should come at the training stage, which is consistent with what I described above.

Finally, visual art is one of the most primitive forms of human communication and delivers unspeakable kinds of emotions, ideas and world interpretations. It has the ability to connect individuals and bridge cultures, and should not be washed away by machine mimicry. I hope the Copyright Office can do as much as possible to help us preserve its value and protect human artists.