To Whom It May Concern at the US Copyright Office,

My name is Brian Penny, and I'm a freelance ghostwriter and journalist. I also have an AI stock image portfolio of over 10,000 images currently selling raw MidJourney outputs on marketplaces like Adobe Stock. I am familiar with the technologies and previously wrote your office about the copyright on Kris Kashtanova's AI comic book "Zarya of the Dawn." I also signed an open letter to protect human artists to Senator Chuck Schumer.

I published Midjourney's Discord server's online beginning, and I think it's important you read. It's very clear as early as February 2022 that David Holz and Emad Mostaque (who later released Stable Diffusion) know how copyrights work. In fact in April 2022, they spoke to Brad Templeton, chairman emeritus of the Electronic Frontier Foundation (EFF) about the copyright decision on his brother's Dr Seuss/Star Trek mashup.

The screenshots are illuminating, and you can find it yourself by logging into the Midjourney Discord, search the phrase "You are going to have a copyright problem." And sort the results by "Old."

I am writing you today in response to your call for comments on generative AI rules. I have a lot of thoughts about this and will be addressing all of your questions in detail. The top-level idea I want you to understand is that existing copyright law is perfectly fine the way it is. I believe the only thing necessary is to clarify how it relates to the inputs and outputs of generative AI, which I'll discuss in this letter.

The one change I would suggest in the modern age is to enforce copyright even if we do not register it. Two perfect examples are social media and livestreaming—artists often create works of art during a livestream, and there have been problems. In October 2022, art streamer ato1004fd live streamed a Genshin Impact art piece and a user fed the in-progress work into an AI generator to complete it before him. In these cases, the Copyright Office must maintain the spirit of the law and allow the human artist to hold copyright (by right of his livestream) over the person who watched, finished it, and could potentially file a copyright first. This often occurs on social media platforms like Twitch, Twitter, Facebook, and TikTok, and our copyrights must be upheld the instant we make them public, regardless of whether we file officially with you. There must be an exception that allows us to pursue legal action to protect our work, even if we have not registered the work with you.

To answer your questions:

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?

I am not anti-technology. I think artificial intelligence is an important technology that has both significant benefits and risks:

Benefits of AI:

Efficiency and Scalability: AI can rapidly generate content, making it useful for various applications like data journalism, data analysis, and entertainment purposes.

Creativity Support: Artists, writers, and other creative professionals can use AI to enhance their creative process, for example, in generating ideas for titles and subheads or changing the style of existing work they already created.

Risks of AI:

Economic Displacement: Freelancers and content creators risk being further exploited, as industry marketing makes their potential clients believe AI could produce similar work more quickly and possibly at a lower cost. This is a fallacy, however, as AI cannot match the quality of a skilled artist, writer, or musician. Examples in the public space (such as Fable Simulation's South Park Simulation) greatly exaggerate the capabilities of AI. A lot of uncredited human work went into producing that faux-AI South Park episode. This misrepresents the technology's capabilities while devaluing the human creativity and effort from unnamed humans that went into creating it.

Intellectual Property Concerns: Using copyrighted materials to train these models creates ethical and legal complexities around who owns the generated output. AI companies have not been transparent about how their models are trained. They have not taken any steps to respect copyright and continue to run billion-dollar companies in the guise of research labs. It is not academic research taking place, but instead widespread copyright and trademark violations. Your office is pivotal to stopping this horrendous crime.

Artistic Authenticity: AI-generated work is already too prevalent. It is infesting every marketplace and platform, and it is not being properly labeled. We have seen this play out in a variety of nefarious ways, including artists having to compete with AI replicas of themselves. When artists complain, they have LoRa models trained on their work (or even images of their real-life bodies) to further troll and humiliate them. Commercial copyright and personal human rights share many similarities in their need for protection.

In line with my beliefs in the four pillars of Consent, Compensation, Credit, and Transparency, I think the USCO must clarify existing copyright law as it pertains to AI, to protect the interests of all stakeholders involved, which includes the creators of the input data and the users creating outputs. The ambiguity created a gray market where everyone on both sides is afraid because of unscrupulous AI companies. As an illustration of this, consider how Kris Kashtanova did not properly reveal she used AI in the creation of her comic, causing your office to amend the copyright.

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?

I'm a ghostwriter. If ever there were a job in danger of being replaced by AI, I must believe writing things others take credit for is at the top. But I also blog and do investigative journalism under my name, having been published in outlets like New Times, Forbes, Paste Magazine, High Times, and Cracked. In the context of journalism and blogging, the increasing use of AI-generated material does raise several unique issues:

Unique Issues:

Erosion of Trust: AI can undermine the public's trust in news, making it harder to discern between human-analyzed reporting and algorithmically generated content. Google's AI chatbot Bard made a factual error in the launch marketing that dropped $100 billion from Alphabet's stock. Gannett paused its AI-generated articles after going viral on social media for publishing a ridiculous mistake a human would never make. CNET also received backlash for publishing incorrect information because of using AI, and it was not always immediately clear AI was used, as it wasn't always correctly disclosed by these outlets.

Ethical Reporting: AI models lack the ethical considerations that human journalists adhere to, such as bias avoidance, accurate sourcing, and the potential consequences of publishing sensitive or controversial information. In addition, AI can only be used to summarize data—it cannot speak to the multiple sources of information needed for proper reporting.

Bylines and Accountability: AI-generated articles lack a human byline, making it difficult to hold anyone accountable for inaccuracies or ethical lapses in the content. The anthropomorphized media and marketing buzz around AGI and other fictional forms of technology attempt to shift the blame from humans to the inanimate tools they are using. This is a key point the USCO should consider, as accountability is important for copyright to function as designed.

Intellectual Property: Journalists could see their original reporting being used as training data without consent, compensation, or credit, violating what I consider to be the four pillars of ethical data usage. Large news publishers, from Disney to the New York Times and NPR, are already blocking AI data scraping bots and starting legal actions against manufacturers of large-language models (LLMs).

Writers across the board face an uphill battle to protect our work, and real journalism is absolutely essential to our society.

Compared to Other Sectors:

Speed Matters: In journalism, the speed at which news is broken is crucial, making the efficiency of AI both a potential advantage and a risk in terms of accuracy and ethics. Social media already causes problems with protecting the CMI data associated with protected works, and it's difficult for media outlets to enforce any type of action against infringement.

Higher Ethical Standards: Journalism has a unique set of ethical considerations compared to other sectors. AIs do not learn like humans and cannot understand these ethics--they harm the profession's integrity. Real reporting requires speaking to various first-hand sources to learn exactly what happened, and AI is designed to only regurgitate derivatives of data already input into it. It will just make up quotes. It cannot tell you what is happening right now and, therefore, has limited potential use cases in journalism besides header/subhead ideation and content summarization.

Public Interest: Unlike other creative sectors, journalism is critical in democratic societies. Misuse of AI in this sector could have societal implications beyond economic or individual concerns. We are entering an election year very quickly, and AI-generated deepfake images, videos, voices, and text will all play a significant role (if they aren't already).

The USCO needs to account for these sector-specific challenges when considering how copyright laws should be clarified and applied in the context of AI.

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.

On the Danger of Stochastic Parrots by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell

Towards a Standard for Identifying and Managing Bias in Artificial Intelligence (NIST Special Publication 1270) by Reva Schwartz, Apostol Vassilev, Kristen Greene, Lori Perine, Andrew Burt, Patrick Hall

Please speak to Bender, Gebru, and Mitchell. They are experts in ethical AI and have been thus far excluded from the conversation. They know what they are talking about, and their opinions should be taken with a much higher weight than CEOs like Sam Altman at OpenAI or legal representatives like Ben Brooks at Stability AI.

4. Are there any statutory or regulatory approaches that have been adopted or are under consideration in other countries related to copyright and AI that should be considered or avoided in the United States?

How important is international consistency in this area across borders?

China has one of the best approaches so far, ensuring the AI is not discriminatory and mandating that AI outputs are correctly labeled and disclosed as AI. You're doing it right if credit, compensation, consent, and transparency are honored. Thanks to the Berne Convention, you will be setting a standard that's followed by most countries, unless you drop the ball and allow other nations to pick up the slack.

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.

I believe new legislation specifically tailored to address copyright or related issues with generative AI is not necessarily needed. The existing copyright framework is generally robust; what's needed is clarification on how this framework applies to the realm of AI.

Here are some areas that might benefit from legislative consideration, given the unique challenges posed by AI:

Data Usage Ethics: Legislation could mandate the four pillars of **Consent, Compensation, Credit, and Transparency** for companies that collect data for training AI models. This could include requiring explicit opt-in consent from content creators and providing mechanisms for fair compensation and credit where it's due.

Accountability: New clarifications could outline who is legally accountable for AI-generated content (inputs and outputs). This would be especially important in sectors like journalism, where accountability is crucial. Still, it will remain important in determining who is responsible if I prompt an image "in the style of (artist name)," let alone whether they get compensation/credit for their work being used without their name.

I believe new legislation specifically tailored to address copyright or related issues with generative AI is not necessarily needed. The existing copyright framework is generally robust; what's needed is clarification on how this framework applies to the realm of AI. Some suggestions:

Public Disclosure: Laws should require companies and publications to indicate when AI has generated content, aiding public awareness and accountability. Far too many instances of AI-generated content being passed off as human-made exist, causing major liability issues as authors risk losing copyrights to books over covers, misinformation is spread, etc. Check out this case with a flood of AI-generated mushroom foraging books on Amazon, often containing incorrect information that can kill you.

AI-Generated Work Registry: A public or restricted-access database should be created to store and identify AI-generated work, helping with the issue of copyright enforcement. This should be a separate section from the rest of the USCO's archive at the Library of Congress so that we do not lose our own human culture and history to the AI.

Royalty and Licensing Schemes: Given the potential for wide-scale distribution of AI-generated works, a new royalty and licensing framework could be beneficial for ensuring that original copyright holders are compensated based on how their work was used. If AI "learns like a human," then it needs to take student loans to pay six figures or more for its education like humans do at that level. Computers are binary, and it is absolutely possible to trace the 1s and 0s to know how much of each person's work was used to create a new work. These percentages should be compared.

For example, if I prompt Midjourney for "horse" and get under 2% of any given work versus prompting "horse in the style of Sarah Andersen" which would weight it as 10% of several of that artists work, versus training a LoRA model specifically on her work and raising that level to 20% or more. These percentages need to be understood and applied to the framework of the law, thanks to AI's proliferation.

While the existing copyright law framework is strong, these additional layers could help to address the unique complexities that AI brings to the table. These guidelines would aim to preserve the integrity of various industries while still allowing room for technological innovation.

Training

If your comment applies only to a specific subset of AI technologies, please make that clear.

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

I will be addressing all forms of training (voice, images, writing, etc) at all stages (scraping, pre-training, training, human-reinforced learning). However, this is all limited specifically to generative AI and should not be applied to other forms of AI, such as cameras that rely on AI facial recognition, autozoom, and more.

First let's break down the types of materials.

Types of Copyright-Protected Training Materials

Text Data: This can include articles, blog posts, books, guides, scripts, books, social media posts, and other forms of written content.

Images and Art: This includes photographs, illustrations, vectors, typography, and other visual arts.

Audio Data: This includes music, podcasts, radio, social media, and other types of recorded audio.

Video Content: This includes films, TV shows, livestreams, shorts, and other video materials.

Code and Software: While not always considered in the same category as other forms of art, software can be copyrighted and is also used in training. Because of the nature of software, AI outputs are very likely to be infringed.

Interactive Media: This includes games, apps, websites, and other types of interactive content.

How These Materials Are Collected

Web Scraping: Data is gathered via automatic collection of data from various online sources. Often the party scraping the data is not the same as the party training its algorithms with it.

Public Datasets: Public datasets contain copyright-protected material that was included without proper authorization or understanding of the implications. “Public domain” as recognized by the USCO is not the same as “publicly available” as stated by AI companies, and this distinction must be made.

User Contributions: In some models, users contribute data which could potentially include copyrighted materials. All datasets should be opt-in only.

Licensed Data: Ideally, some organizations obtain licenses to use copyright-protected materials, although this is not always the case. For example, OpenAI signed a licensing deal with the Associated Press to use its data for its LLMs, but it did not sign the same deals with the NY Times and other publishers now considering lawsuits for copyright infringement. Getty licensed its image content to Bria AI and Shutterstock licensed to OpenAI. These deals are key to the survival of creative works in the age of AI.

Content Curation

Filtering: Datasets are usually filtered to remove irrelevant information and address bias. The LAION 5B dataset, for example, was trimmed from 5 billion images to 2 billion images for LAION 2B aesthetic. This is also true of LLMs and other forms of generative AI. However, none of them appear to have filtered for sensitive information (see medical records found in LAION 5B), and many companies have banned employees from inputting proprietary information into these machines (see Samsung).

Data Augmentation: The existing data is then modified or expanded to create a more robust dataset. This includes rotation and scaling to augment the dataset, use of synthetic data, textual variations, and data fusion, in which multiple sources are fused together. This augmentation is key to Adobe’s Firefly AI success, as mentioned during their Senate Judiciary Committee testimony.

Quality Check: Human curators go through the datasets to ensure the quality and relevance of the data, alongside automated quality assurance. During this stage, it’s vital that steps are taken to address systemic biases on top of what should have been done in previous steps.

Special Considerations

Consent, Compensation, Credit, and Transparency: Given my strong belief in these four pillars, the collection and curation process should adhere to these principles. It must be opt-in (consent), creators must be paid (compensation), and they should have their CMI attached (credit). The only way we can ensure this is through transparency.

Post-Training Reinforced Learning: RLHF stands for Reinforcement Learning from Human Feedback. In the context of machine learning and AI, RLHF is a technique that uses human feedback to train a reinforcement learning model. The idea is to improve the model's performance by incorporating insights gained from human interactions, which could include reward signals, corrections, or more explicit forms of guidance.

For example, an AI system trained using RLHF could start with a preliminary model based on existing data and simulations. Humans then interact with the model, providing feedback to create a reward model for reinforcement learning. The AI system is then fine-tuned based on this reward model to improve its performance in the tasks it was designed for. People must be compensated, credited, and asked for explicit written consent for this RLHF training the same way they are for pre-training and training.

The curation process is not just a technical necessity; it's a crucial stage where ethical and legal guidelines must be rigorously applied to ensure that the AI model is effective and responsible. copyright law needs to provide clear guidance on what forms of data collection and curation are acceptable and what forms violate the principles of consent, compensation, credit, and transparency.

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)?

Developers acquire materials or datasets for training AI models from a variety of sources, and the extent to which these materials are first collected by third-party entities can vary widely. Here's a breakdown:

Public Datasets: Many developers start by using publicly available datasets, often curated and published by academic researchers or governmental bodies. These datasets may be specialized for various purposes such as natural language processing, computer vision, or other tasks. They often contain copyrighted work that was scraped off the internet without permission. These academic datasets should not be allowed to be used commercially. If they are, they cannot contain copyrighted material that's unethically scraped without consent, compensation, and credit.

Web Scraping: This method involves automated collection of data from websites and online sources. It's a common practice but is fraught with legal and ethical complications, particularly around consent and copyright. This occurs when the company decides not to use existing data sets (which often were also scraped) in favor of creating their own. This is not often used, as AI companies prefer to make liability difficult to track (ie. “I didn’t do it—I just used a publicly available model!” while stealing).

Purchased or Licensed Data: Commercial data providers offer specialized datasets for a fee. These datasets may be collected and curated by the providers themselves or acquired from third-party sources.

In-House Data Collection: Some organizations collect their own data for training purposes. This could include customer interactions, sensor data, or other forms of data relevant to the business. It’s typically done during the RLHF.

User-Generated Content: Platforms with large user bases often use data generated by their users to improve their AI models, although this often raises ethical and privacy concerns. We see this especially in social platforms like Facebook and Twitter/X, and it's important that people's right to their data is respected the instant they publish it to one of these platforms, as their TOS is very restrictive. I should not have to register a Facebook post or tweet with you to own it.

Academic Collaborations: Universities and research institutions often partner with businesses to share data, subject to academic ethical review boards and data usage agreements. Again, this must be done in a way that respects copyright—if MIT makes a model on Mickey Mouse, it must get consent and provide compensation to Disney for commercializing its academic research. Being a university or research lab

Third-Party APIs: Some developers use APIs to gather data from third-party services, which may include social media platforms, financial data feeds, or other specialized data services.

In terms of ethical considerations, regardless of the source, it's crucial to adhere to principles like Consent, Compensation, Credit, and Transparency, especially when using third-party collected data, to ensure that data is being used responsibly.

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.

Training materials are processed and encoded into a format that can be used by the AI model during the training phase. The nature of the reproduction could vary based on the type of data—text might be tokenized, images might be converted into numerical matrices, and so on. The reproduction is often temporary but could persist in storage for the duration of the training process and potentially beyond.

From a copyright perspective, this act of reproduction and use raises serious concerns. It implicates the exclusive rights of copyright owners, particularly if the training materials were used without permission, which is why I strongly advocate for the four pillars of Consent, Compensation, Credit, and Transparency.

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

The inferences gained during training are stored in the model's parameters, which are essentially numerical values that the model adjusts during learning. These parameters shape the model's behavior and decision-making but don't contain the training data in a recognizable or reconstructable form.

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?

Unlearning specific inferences is theoretically possible but often not practically or economically feasible with current technology. The learning in most AI models is diffused across numerous parameters, making it difficult to pinpoint and remove the influence of specific training material. Retraining the model without the contentious data is often the most straightforward way to "unlearn" specific inferences, but it's a resource-intensive process.

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?

It's generally not possible to definitively identify whether an AI model was trained on a specific piece of material without access to the underlying dataset. Even if the model generates output that resembles a known piece of copyrighted material, that similarity could be coincidental or arise from the model's exposure to similar data types, rather than that specific piece.

Given these complexities, I maintain that clear ethical and legal frameworks are crucial for guiding the use of copyrighted materials in AI training processes.

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?

Considering the Google v. Oracle America and Andy Warhol Foundation v. Goldsmith decisions, evaluating the "purpose and character" of using copyrighted works to train an AI model becomes more nuanced.

Relevant Use to Be Analyzed

The relevant use to be considered should be the ultimate application of the AI model. Is it commercial or non-commercial? Does it substantially affect the market for the original work? Does it create new opportunities for creativity, much like Google's reimplementation of Java APIs did? Again, I must reiterate that if I'm using it for academic purposes and give it to you to sell, that sub-license does not magically make the commercial use academic. This switcheroo from the AI industry ***MUST*** be stopped under all circumstances, as it spits directly in the face of the spirit of copyright law. Midjourney, Stability AI, OpenAI, et al do not get to circumvent copyright law just by calling themselves research labs. A school that steals is still stealing.

Pre-training and Fine-tuning

Different stages like pre-training and fine-tuning could indeed raise distinct considerations under the first fair use factor. Pre-training, which usually involves generalized learning from a broad dataset, is fully derivative as it depends entirely on the data given to it. Fine-tuning, on the other hand, usually narrows down the model's capabilities and might be more aligned with the original purpose of the copyrighted material. The closer the fine-tuning aligns with the original copyrighted work, the less likely it is to be considered transformative. Even worse, when an person's work is invoked by a user (ie "dinosaur in the style of Greg Rutkowsky"), it is violating their right to publicity while raising the percentage of their work included in the output.

Both pre-training and fine-tuning must be closely examined, as each has a large effect on the output. ***It's also important to understand that there are two models involved—the generation model and the CLiP model. If the CLiP model can understand copyrighted work (using author names "book in the style of***

Stephen King" or work names "Afghan Girl") then it can recreate it even if it's not included in the generation model!

Google v. Oracle America

Google v. Oracle sets a precedent for viewing transformative use expansively, but the ruling particularly noted the value added by Google's new implementation and the new opportunities it created for third-party developers. If an AI company could show that its model enables significant new forms of creativity or utility, that could weigh in favor of fair use. However, this will prove impossible as ***<u>all AI outputs</u>*** are ***<u>always</u>*** derivative of the input works. They never make up any novel nor new data, instead making composites using small percentages of many images. Infringing on large volumes of work should not negate the infringement. There is no world where it's fair to justify theft by scaling it—if it were possible, I'd be robbing every bank right now.

Andy Warhol Foundation v. Goldsmith

This case sets a more stringent standard for what counts as transformative, focusing not just on aesthetic differences but also on "distinguishably different artistic purposes." AI companies using copyrighted material for training will need to consider whether their use serves a new or different purpose than the original work, to argue transformative use under fair use criteria. As we've seen with AI outputs being submitted for copyright (thus used for commercial purposes), artists, writers, authors, musicians, voice actors, etc. are being forced to compete with AI-generated imitations of their work in search engines, marketplaces, and social media platforms.

Considering these legal landscapes, I maintain that copyright law needs only clarification when it comes to AI, emphasizing on the four pillars: Consent, Compensation, Credit, and Transparency. This will ensure that the spirit of existing laws apply to these new technologies.

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?

Collecting and distributing copyrighted material without properly licensing it is a crime. Both public libraries and schools license all their content. A school band director can't plan a public performance of a Taylor Swift song without obtaining the license. This should apply online as well. I can't broadcast NFL games without the express written permission of the NFL. I can't access Netflix to "learn" what's in their archive without paying for a membership license. And the TOS of these licenses are set by the copyright owners who determine how their work can be used. If people don't want to be included in training, whether academic or not, they shouldn't be. And "publicly available" is not the same thing as "public domain." These important differences ***<u>MUST</u>*** be enforced, and there should NOT be any exceptions for the data scraper or the model trainer.

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.[44]

How should the fair use analysis apply if AI models or datasets are later adapted for use of a commercial nature? [45]

Does it make a difference if funding for these noncommercial or research uses is provided by for-profit developers of AI systems?

Here is what to consider:

1. Switcheroo

If an AI model initially trained for noncommercial or research purposes is later adapted for commercial use, the "purpose and character" of the use could shift. This should weaken the argument for fair use since commercial uses are generally less favored in a fair use analysis. In short, if I set up a bookstore in the parking lot of the public library and begin selling their books, I'm a thief.

2. Funding

The source of funding can also influence the fair use assessment. If a noncommercial research project is funded by for-profit entities with the implicit understanding that the resulting AI models will eventually be used for commercial purposes, this should weigh against a fair use claim from the outset. We saw this with companies like MidJourney, Stability AI, and OpenAI labeling themselves as research labs to pretend they are performing academic research. Today, OpenAI is on track to earn $1 billion in revenue this year, with MidJourney earning $200 million. These are not academic research labs—they're commercial enterprises making a LOT of money from stolen IP.

The alignment of interests between the research and a for-profit entity could indicate a commercial intent that may weigh against a fair use justification. It could also highlight questions around whether the use is genuinely noncommercial or if it serves as a "test phase" for subsequent commercial exploitation.

While noncommercial or research uses of copyrighted material in AI training are more likely to be considered fair use, the shift to a commercial model or the influence of for-profit funding should substantially weaken such a claim.

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?

Quantity

The volume of training materials used to train generative AI models can vary significantly, ranging from thousands to billions of data points, depending on the complexity of the task and the model. However, this scale should not be an excuse to bypass existing IP protections—copyright exceptions are typically made for using a lower amount of someone's work (ie—remixing and sampling). AI scales this to exponential proportions that need to be considered.

Impact on Fair Use

The volume of copyrighted material used is a consideration in the fair use analysis for copyright exception. A high volume might weigh against fair use, especially if it is a substantial portion of the copyrighted work. However, the "amount and substantiality" are considered relative to the purpose of the use. Because AI models depend 100% on the data input into them, the copyrights on that data should hold.

It's again important to point out that all forms of AI use two different models at least. One model contains the matches needed to match a user's text input as a token to the tokens attached to the generation model. The generation model contains all the work that weighted the models. So, when you type the word "pizza," for example, an initial model (CLIP for images) matches it to the pool of images/text/music/video labeled “pizza,” while the secondary model produces a composite of the best data for pizza.

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?

Measuring the impact on market value is complex. One method might be to assess whether the AI system's outputs are a substitute for or competing in any way with copyrighted work. If they are, this could significantly impact the market for copyrighted material, weighing against fair use. To prove this has already been violated, we can look at the Google image search results for artist Kelly McKernan (currently involved in a copyright infringement lawsuit against Midjourney and Stability AI) to see that she is competing with AI-generated images using her name. This puts her in direct competition with AI outputs in the marketplace.

Scope of Inquiry

The scope of the inquiry could vary:

Specific Copyrighted Work: If the AI output directly competes with a single copyrighted work, that would be a strong indication against fair use. For example, using the word “Afghan Girl” is banned in Midjourney as it is found to be overfitted to the famous photograph of Sharbat Gula taken by Steve McCurry in Pakistan in 1984. This in an obvious copyright infringement using the naked eye.

Body of Works of the Same Author: If the AI output competes more broadly with an author's entire portfolio, this could also weigh against fair use, though the impact may be more diffuse and harder to quantify. In the case above, McKernan’s entire body of work is competing against AI. This is true of many artists, including Artgerm, Greg Rutkowski, and more. Media outlets have also been harmed by having their entire archives used to train LLMs like ChatGPT on what is “true” and “factual” versus “misinformation.” If the AI developers can’t do this without stealing work, then they shouldn’t be doing it until they can afford a proper license from the copyright owners.

General Class of Works: If the AI output competes with a general class of works (e.g., news articles, poems), it might still have a negative impact, but proving this could be challenging because of the diversity within such a class. It’s vital that the USCO encompasses both inputs and outputs to address this.

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)?

Opt-in is the only logical choice, as it aligns with our general data privacy rights. If a company made a new social media platform and added profiles for all of us without our permission unless we opt out, there would be riots. It is a violation of basic privacy rights, and it is obviously wrong that a company could use your name, headshot, and other public information to build a profile that looks like you’re behind it. This is fraud and should not be allowed. Adding the word “artificial intelligence” to this equation does not change the fundamentals.

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

It should be for all use cases. I have to opt-in to donate my dead body to science as a cadaver. I have to opt-in to donate my organs if I die. It should always be opt-in, and I am not aware of any examples in which requiring consent from the owner has stifled innovation in any way. No scientific discovery in human history has ever been slowed or stopped in any way by the need to obtain permission.

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?

Digital rights management (DRM) and metadata can address this. However, I don't believe "opt out" is an affective approach. Opt-out is exploitative. In many cases, we are unaware of the data scraping and usage for training until it is too late. None of the AI companies gave any advanced notice of what was happening, and therefore they were able to train their models without our permission.

According to Spawning AI (hosted on Hugging Face) over 80 million works were opted out of AI training as of March 7, 2023. As of September 15, 2023, their website says that number is 1.4 billion (28% of the LAION 5B dataset). This is unacceptable to have nearly 1/3 of people choose to opt out, and even worse, Stability AI, Adobe, OpenAI, and others have openly admitted they don't know how to untrain models yet. If they can't undo the damage, opt-in is the only ethical, fair, and just way.

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?

Yes it is feasible—every day an estimated 3.7 million videos are uploaded to YouTube, 4.75 billion Facebook posts are published, and 34 million TikToks are uploaded. There are automated systems in place to monitor everything—if you don't believe it, try uploading this Taylor Swift video to any of those platforms and see how far you get before it's removed.

The term OpenAI has 31,300 news articles listed in Google News. The company has 2.5 million Twitter followers, 72,000 followers on Facebook, and SimilarWeb ranks it as the 28th most visited website on the entire internet globally, drawing over 1.4 billion visits per month.

There's no reason OpenAI can't get opt-in consent. I opt into cookie policies on every single website I ever visit, thanks to GDPR (not even a US law to protect us, so thankful to Europe for caring more than we do). I opt in to Apple and Google and Microsoft and Facebook services all the time. There's no company too big for me to voluntarily opt in. You don't automatically get a Gmail account unless you opt out. You don't automatically get a Facebook page unless you opt out. You don't automatically get an iPhone unless you opt out. There's nowhere in the entire world where opt-out has ever been necessary, and we've progressed quite a bit as a species over the millennia despite that.

Opt In is the only ethical way.

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?

If an objection is not honored, various remedies could be available depending on the specific circumstances and jurisdictions involved. Here are some options:

Existing Remedies for Infringement

Cease and Desist Orders: A legal injunction could require the AI developer to stop using the copyrighted work immediately.

Monetary Damages: Compensation could be sought for any financial loss incurred due to the unauthorized use of copyrighted material.

Statutory Damages: Some jurisdictions allow for pre-determined statutory damages for each act of infringement, regardless of actual loss.

These existing remedies could be appropriate for straightforward cases of copyright infringement. However, the unique challenges posed by AI could warrant additional considerations.

Separate Cause of Action

Specific AI-related Remedies: This could include actions to 'de-train' the AI model from the copyrighted material or specific penalties for using copyrighted material in AI training without honoring an opt-out.

Fine-Grained Penalties: A separate cause of action could allow for more nuanced penalties based on the extent of the copyrighted material's use, its impact on the AI model, and the potential market impact on the original work.

Additional Measures

Transparency Requirements: Failure to honor an opt-out could also trigger requirements for public disclosure, adding a reputational cost to legal penalties.

Third-Party Arbitration: An independent body could assess whether an opt-out should have been honored and what remedies are appropriate, providing a quicker resolution than court systems.

Given the novel issues presented by AI, a separate cause of action tailored to address these unique challenges may be warranted. It could exist alongside traditional copyright infringement remedies, offering a more nuanced approach to resolving disputes in this emerging field. Separating AI from human-generated work could be necessary to preserve our human culture as well.

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?

In cases where the human creator does not own the copyright, the question of whether they should have a right to object to an AI model being trained on their work presents a complex legal and ethical challenge.

Moral Rights: The Berne Convention provides creators with "moral rights" that exist separately from copyright and allow them to object to uses of their works that they find objectionable, even if they do not own the copyright. Such rights are less recognized in U.S. law but do exist in various forms in other countries due to this international agreement. The USCO should refer to the Berne Convention while forming your rules.

Attribution: The human creator has a default vested interest in how their work is being used, especially if the AI model's output could be seen as a reflection or derivative of their original creation. This is why a default “Opt Out” status should be placed on all works not in the public domain the instant they are created. Functionally, this would mean including an opt-in within metadata or robots.txt to allow web scraping. This is not much different than the current DRM environment in which The New York Times, for example, hides information behind a paywall. These paywalls are bypassed by archiving the page, however, allowing piracy to run rampant. Your decisions on generative AI will also affect applications like search engines and the internet archive.

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

It can be obtained the same way YouTube, Netflix, Spotify, Facebook, Twitter, Apple, and everybody else obtain it. I consent to the TOS of their services when I use them. This is a very simple concept that exists everywhere we look. There is no need to reinvent the wheel just because we’re using generative AI. There is absolutely no reason the current opt in system we’ve been using this entire time can’t still work. I opt in to cookie tracking on every single website I visit. We have laws for third-party cookies. We have laws for data privacy. These apply to generative AI—why would they not? Why are we acting like this isn’t the way things have always been? It’s insanity to fall for industry marketing like this. Nothing changes.

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

Of course—why would it not be? Spotify successfully licensed its archive of over 100 million songs, 5 million podcast titles, and 350,000 audiobooks. YouTube successfully licensed its archive of over 800 million videos. Twitter successfully licensed 500 million tweets per day by August 2013. Shutterstock successfully licensed over 724 million images. Nobody has ever had any issues licensing at scale in human history—there are existing systems, and none of this magically changes with the introduction of AI.

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?

Yes of course!

See the entire internet—try ASCAP and BMI for music. Upload this Justin Bieber music video to Facebook, YouTube, Twitter, or anywhere else—watch how fast it’s flagged for copyright infringement before it ever even allows you to upload it. These systems all exist already and have for a long time. There’s no reason to disrupt it and fundamentally change everybody’s business just because AI companies don’t want to adhere to the same rules and tools that everybody else has always used.

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?

Congress already has a licensing regime in the form of the US Copyright Office as a subset of its Library of Congress. Rates can be negotiated by the two parties involved in the contract just like they always have in contract law. There's no reason for Congress to step in any further besides to enforce current copyright laws on AI companies that are not adhering to.

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

No-- An Extended Collective Licensing (ECL) scheme allows a collective management organization to license rights on behalf of all rights holders in a particular category unless they opt-out. This can be done by existing organizations, such as we're seeing play out with the collective bargaining between Hollywood writers, actors, and studios. There are already groups set up for collective bargaining on behalf of writers, authors, actors, and more. Musicians and visual artists could definitely use an advocacy group to help with unionizing, but I'm not sure Congressional intervention is the best way to handle that.

Congress may possibly consider setting laws to protect freelancers and gig workers doing work-for-hire across all industries (creative or not). But in the realm of copyright, this type of collective bargaining is unlikely to help the little guy the way it's meant to.

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

If you go that route, then no. It should be evenly applied to all disciplines. If it can't be broadly applicable, then it's unlikely to be the correct path forward.

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)?

Scope of License: The license needs to clearly define what kinds of uses are permitted, including derivative uses, modifications, and scope of distribution.

Attribution and Metadata: Maintaining accurate metadata for millions of pieces of content could be extremely challenging but is necessary for proper attribution and tracking.

Privacy: Some data may have privacy implications that require special handling, such as anonymization.

Interoperability: There could be technical issues with how different datasets, each with its own licensing constraints, can be mixed and used together.

Enforcement and Monitoring: It may be difficult to monitor and enforce compliance with licensing agreements, especially if the datasets are large and widely distributed.

Cost: High licensing fees could stifle innovation and limit access to AI training data for smaller entities.

Jurisdiction: Legal frameworks may differ across countries, posing challenges for international AI projects.

Despite these challenges, we know that companies regularly achieve it. Twitter/X regularly updates its terms of service, and people regularly accept them. It's not hard to get Opt In. The following six parties are responsible for licensing:

Dataset Curator: If a dataset is curated from various sources, the curator may bear the initial responsibility for ensuring all data is appropriately licensed. LAION is an example of dataset curators.

AI Model Developer: The developer should ensure that the training data used complies with all licensing requirements, especially if new data is introduced after initial curation. Stability AI is an example of an AI model developer.

Employing Company: If an organization is deploying an AI model, it should conduct due diligence to ensure that all licensing requirements have been met, as it could be held responsible for any infringement. Poe from Quora is an example of an employing company.

Model Hosts: AI models are typically hosted in centralized places, such as Hugging Face and Civit AI. These hosts should be held responsible for performing due diligence to ensure they aren't hosting infringing or otherwise illegal content.

Aftermarket Models: Refined models (such as LoRAs for image generators) can be created and traded to teach the base model new things. These models should be properly licensed to avoid infringement, as they drastically raise the percentage of an author's work used in the output.

End User: End users have the capability of creating a book in the style of CNET or a painting in the style of Karla Ortiz. When invoking an artist's name, Adobe research shows that it greatly increases the percentage of their work being used. This means that the names of authors/artists and their works need to be protected.

Given the complexities and multiple parties involved, clear contractual language and agreements specifying responsibilities can be valuable in navigating these issues.

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.

Yes – computers are binary at their base. It's all either a 1 or a 0. And even "random" number generators are not actually random. They're based on a seed, and if you know the seed, you can easily break encryption, for example. For AI companies to say it's impossible is laughable—look at Adobe's research showing percentages. It's an awful company, but it disproves the false claims from "research labs" like OpenAI and Stability AI.

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

Proper licensing would provide a huge boon to the economy, as we have a lot of content available. It would provide revenue for copyright holders and resolve the legal uncertainty surrounding these models. Companies like OpenAI and Midjourney earn huge profits. OpenAI is a subsidiary of Microsoft and has the money to pay for licensing. Midjourney earns $200 million a year and grew without VC funding. They can all afford to license, and if they can't—that's life. I can't afford to license Disney movies, but that hasn't stopped Netflix, Amazon, Hulu, Max, Peacock, Roku, and so many others from existing. If anything, we have too many streaming services, despite the costs associated. It's impossible that AI companies can't scale without stealing work. If they need money to pay for licenses, then Congress can enforce existing copyright laws and then pass another law subsidizing generative AI so that you can compete with China or whatever. I think it's unnecessary, but it would resolve the problem AI companies convinced you exists.

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

Yes—your decision on Kris Kashtanova's "Rose Enigma" copyright application will decide this and must carefully tread, as AI means we can track things down to trace percentages being used in different works. As stated numerous times above, it is vital that we respect current copyrights and not allow AI companies to bypass them nor change the rules to suit them. The current system worked this long, so there's no reason it still can't work today.

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?

Implementing a requirement for developers of AI models and creators of training datasets to collect, retain, and disclose records on the materials used for training would be a monumental step toward transparency and fairness. Such transparency would empower copyright owners, providing them the clarity they need to understand how their works are being utilized, whether for commercial or non-commercial purposes.

This approach could also streamline the process of licensing, as copyright owners would have a clearer understanding of how their works are being employed, potentially leading to more equitable licensing agreements. In cases where unauthorized use is occurring, having a transparent record would enable copyright owners to take appropriate legal actions more efficiently.

So, how could this be implemented?

Metadata Tagging: Developers could integrate metadata tagging systems to mark copyrighted materials as they are fed into the training algorithms, allowing for automated tracking.

Secure Record-Keeping: Encrypted databases could be maintained to ensure secure storage of records, accessible only to authorized parties.

Regular Auditing: An independent third-party could be mandated to conduct regular audits of these records to ensure accuracy and compliance.

Public Disclosure: A simplified, non-technical summary of the records could be made publicly available, offering copyright owners an easy way to ascertain if their material has been utilized.

Real-Time Notifications: An API could be created that allows copyright owners to receive real-time notifications whenever their material is added to a training dataset.

Standardized Formats: To make this scalable, a standardized data format for recording the usage of copyrighted material could be implemented.

By pursuing this level of transparency, the AI community can build trust, not only among developers but also with content creators and copyright owners. The complexities surrounding AI and copyrighted material will only grow, and proactively building a transparent infrastructure can serve as a strong foundation for the ethical and fair use of intellectual property in AI

15.1. What level of specificity should be required?

The level of specificity required should strike a balance between giving copyright owners enough information to understand how their works are used and maintaining the efficiency and viability of the AI development process.

Work Identifier: A unique identifier for each copyrighted work used should be a minimum requirement. This could be as simple as a title, ISBN, or other standard identifier, which would then be tagged and logged.

Type of Usage: It would also be beneficial to specify the type of usage—whether the material is used for training, fine-tuning, or validation, for example.

Amount of Material: Given that the volume of material used can impact fair use analysis, this too should be logged. The range could be as broad as specifying whether the entire work or only snippets are used.

Commercial or Non-Commercial: The intended use of the AI model could be specified — for example, whether it's for commercial, research, or educational purposes.

Algorithmic Context: Though perhaps more technical, outlining the broad algorithmic context in which the material was used could offer a better understanding of the work's contribution to the AI model. This could be particularly pertinent if the model performs tasks closely related to the copyrighted work.

Being assertively committed to transparency would entail implementing these details at a high level of specificity. This will enable copyright owners to have a comprehensive understanding of how their intellectual property is being used, thereby enhancing trust and fostering a fairer landscape for the use of copyrighted materials in AI.

15.2. To whom should disclosures be made?

Disclosures should be made available to copyright owners, legal representatives, and potentially even to a governing or oversight body specialized in intellectual property and AI ethics. If these disclosures aren't made, it should be assumed that illicit data is used.

Copyright Owners: The primary recipients of these disclosures should be the copyright owners. They have the most direct stake in how their work is used. If Opt-In is respected, this will be easy.

Legal Representatives: In cases where the copyright owner is represented by legal or management entities, disclosures should be sent to these representatives as well.

Governing Body: To standardize the process and ensure compliance, a specialized governing body could be established to oversee these disclosures. This body could serve as an intermediary between AI developers and copyright owners, providing an additional layer of trust and verification.

Public Archive: Consideration should also be given to making a subset of this information publicly available in a way that respects proprietary and privacy considerations. This could foster an environment of openness and promote public trust, such as how Stability AI open sourced everything but the weights for Stable Diffusion. It should be noted the weights are the most important thing, and what they did is akin to having a nutritional label that doesn't contain the daily value percentages and sugar/sodium counts, etc.

Industry Stakeholders: Companies or individuals who utilize the AI models in their operations will also benefit from these disclosures, as it provides them with assurance regarding the legal safety of their operations. Notion, for example, may be held legally liable for its ChatGPT implementation if not receiving the proper disclosures.

By being assertively committed to transparency and making disclosures available to these groups, we can ensure that copyrighted materials are used in a manner that respects intellectual property laws and contributes positively to the development and application of AI technologies.

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

Developers of AI systems that incorporate models from third parties should bear the responsibility for ensuring that those models are compliant with copyright laws. This means actively seeking to understand the lineage of the training data used for those third-party models and ascertaining whether proper licensing or fair use provisions apply. If the third-party models do not meet the necessary legal standards, developers should either opt for compliant alternatives or work with the third-party provider to remedy the shortcomings.

They should also have an obligation to disclose the use of third-party models to both their users and any relevant oversight bodies. This should be done in a transparent manner, clearly indicating which components of their AI system are built on third-party models and providing access to any available records about the training data and processes of those models.

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?

Implementing a comprehensive recordkeeping system would indeed have financial, technical, and time costs. For developers of AI models, there would be the initial investment in creating or adapting a system that can log and store the requisite information about training data, model versions, and compliance checks. Ongoing maintenance and auditing of this system would also entail costs.

However, the value of such transparency and accountability should not be underestimated. In the long term, a well-implemented recordkeeping system could mitigate legal risks, thereby potentially saving significant costs associated with copyright infringement lawsuits. Transparent recordkeeping can also build trust among consumers and creators, fostering a more collaborative and responsible AI ecosystem.

For creators, knowing that their work is being used responsibly can encourage further innovation and creation, whereas the alternative can stifle creativity due to fears of misuse or uncompensated use. Consumers also stand to benefit from this transparency, as they can make more informed choices about the AI systems they interact with, knowing that those systems respect copyright laws.

In essence, while there are costs associated with implementing and maintaining such a recordkeeping system, the long-term benefits in terms of legal compliance, trust, and innovation make it a necessary investment for the future of responsible AI development.

It’s important to understand that recordkeeping is stored on a low-power hard drive, versus the electricity-guzzling nature of GPUs and TPUs that process AI.

From MakeUseOf: “As a rough guide, here are approximate ranges of how much power is used by each component:

CPU:	55 to 150W
GPU:	25 to 350W
Optical Drive:	15 to 27W
HDD:	0.7 to 9W
RAM:	2 to 5.5W
Case fans:	0.6 to 6W
SSD:	0.6 to 3W”

The costs of storing a recordkeeping system is negligible compared to the computing power in AI. Please contact Dr Sasha Luccione at Hugging Face to discuss this with her, as she is an expert in the environment costs of AI.

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

Notification to copyright is an important safeguard to ensure that their works are being used in a way that aligns with their intentions and legal rights. If opt-in is respected, this should be no problem. A systematic approach to notification would add a layer of accountability that could help in preempting legal disputes and fostering trust between AI developers and copyright owners.

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?

Outside of copyright law, other laws that could require the retention or disclosure of training materials for AI systems. For instance, trade secrets law could potentially intersect with these issues, especially if the training data or the model itself were considered a trade secret. Consumer protection laws might also come into play if the lack of transparency around training data could be shown to mislead users in some way.

Strict criminal laws should also apply in some cases. Consider the flood of Undress Apps on the internet. Searches for these illicit apps exploded since the release of Stabile Diffusion. Their sole purpose is to undress pictures of people without their consent. They are charging people money to make it easier to commit a digital form of sexual harassment. Please refer to My Image My Choice, which has been combatting this problem since it was AI deepfakes (see Atrioc Twitch Deepfake scandal).

Additionally, data protection laws, such as the California Consumer Privacy Act (CCPA) or the General Data Protection Regulation (GDPR) in Europe, may also impose certain requirements for data retention and disclosure, although these laws focus more on personal data rather than copyrighted materials. Data scrapers who make models could be in violation of hacking laws (see Stability AI founder Emad Mostaque's claims of accessing any data, public or private along with Twitter's recent lawsuit against the nonprofit Center for Countering Digital Hatev), Safe Harbor, or even HIPAA.

See Dr Gupta, an AI physician made by convicted felon and pharma bro Martin Shkreli. Although it says it's for entertainment purposes, it does ask users to input sensitive medical information while openly bragging on Twitter that they do not follow HIPAA because they aren't a hospital. Shkreli has received a lifetime ban from the pharmaceutical industry, and his creation of an AI doctor should be considered a violation of a variety of healthcare and human safety laws.

Generative AI Outputs

If your comment applies only to a particular subset of generative AI technologies, please make that clear.

Copyrightability

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

Under current U.S. copyright law, works generated by AI are generally not considered eligible for copyright protection because they are not created by a "human author." However, the extent of human involvement in the generative process can blur these lines. For instance, a human who closely guides the output of a generative AI system through curated training or by providing detailed prompts may be able to claim some form of "joint authorship" with the machine. The U.S. Copyright Office's Compendium of U.S. Copyright Office Practices currently specifies that works must be created by human beings to be copyrightable but does not offer extensive guidelines on situations involving AI-human collaborations.

Factors that may be relevant to determining human authorship in this context could include:

The level of creative control exerted over the AI's training and output: Did the human creator select unique or specialized training materials or provide detailed, specific prompts?

Kashtanova's "Rose Enigma" combines a hand drawn image fed into Stable Diffusion via Image-to-Image, along with a prompt to create the final output. However, the AI model used contains copyrighted content and is infringing upon existing copyrights. I believe that in order for her to receive that copyright approval, she would need to wait until Stability AI and Midjourney properly settles their lawsuits from Getty and artists Sarah Andersen, Karla Ortiz, and Kelly McKernan.

The originality of the contributed elements: Is the human's contribution sufficiently original to merit copyright protection?

Lawsuits such as Open Source Programmers vs GitHub Copilot, Microsoft, and OpenAI have much more direct copyright infringement because of the overfitting nature of functional software programming. In addition, there was recently a book cover discovered to have been made by feeding a collage of multiple artists' work into the same process Kashtanova used. This book, "Imagination Manifesto" by Ruha Benjamin, should lose its copyright protection (even if it is not registered with the USCO per above comments) until the AI and artists/works used in the collage that made the cover image are properly labeled.

Given the evolving landscape of AI and creative works, these questions are ripe for legal exploration and may require legislative or judicial clarification in the years to come, and your decisions here will have long-lasting impacts. Please think through exactly where to draw that line and make that line applicable specifically to generative AI usage so that people know what is right and wrong.

19. Are any revisions to the Copyright Act necessary to clarify the human authorship requirement or to provide additional standards to determine when content including AI-generated material is subject to copyright protection?

As AI continues to evolve and its applications in creative fields grow more complex, the current framework of the U.S. Copyright Act may not adequately address questions of human authorship and AI-generated material. Legislative revisions could provide clarity on these issues. Specifically, amendments could:

Define the Level of Human Involvement: The law could specify the degree of human involvement necessary for a work involving AI to be eligible for copyright. This would need to be relatively vague while also giving examples like those above of the right and wrong way for creatives to file our paperwork.

J**oint Authorship**: The Act could be revised to include explicit provisions about joint authorship between humans and AI systems, outlining the rights and responsibilities of each party. The collaborative nature of humans and their inanimate tools that are incapable of making decisions should be broadly applied—if the rule doesn't make sense when replacing "AI" with "guitar," "paintbrush," or "pen," then it is not the right rule.

Originality Standard: The Act could clarify how the standard of originality applies to works created with the aid of AI. This might include providing criteria to determine what constitutes a "creative" or "original" contribution in the context of AI-generated content. Be sure to consider how the AI auto focus affects a photograph or how auto-tune affects a song. It is vital that any rule changes made to address generative AI applications like ChatGPT and Midjourney do not impact existing works using autotune and autofocus.

Licensing: Consider the complexities of licensing copyrighted material for the purpose of training AI models. Legislators could set clear rules around this to facilitate lawful use while protecting the rights of copyright owners. Opt-in should be mandatory (as it always was) and be clear that compressing data into a dataset is no different than compressing it into a CD or hard drive.

Safe Harbors: The Act could provide "safe harbors" for certain AI activities, especially for true academic research or non-commercial uses, defining the scope and limitations of such protections. Be very clear in these cases, as it's clear that the new growth strategy for tech companies is to pretend they are "research" companies during the first year that's hard for every business. Then they only declare they're a business once forced. Be very clear so that public schools and children have more access to better information than adult organizations (including private universities) trying to make money.

Given the rate of technological change, a flexible legislative approach that can adapt to new advancements would be beneficial. This could include sunset clauses for certain provisions or regular mandated reviews of the legislation's effectiveness in addressing the challenges posed by AI in the copyright space.

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 a nuanced issue that could have various implications for creators, users, and the AI industry at large. On one hand, granting legal protection could provide a framework for monetizing AI-generated content, thereby encouraging investment in research and development.

On the other hand, extending copyright protection to AI-generated material could complicate existing copyright frameworks and potentially stifle human creativity by limiting access to a new class of works. It might also exasperate existing challenges in distinguishing between human-generated and AI-generated content, thus complicating enforcement and dispute resolution.

I also worry about any rulemaking that anthropomorphizes these technologies. It must be made clear that the protections are for the people who are driving the AI inputs and outputs, not the AI itself.

20.1. If you believe protection is desirable, should it be a form of copyright or a separate sui generis right? If the latter, in what respects should protection for AI-generated material differ from copyright?

If protection is deemed desirable, separate sui generis rights might be more appropriate than extending traditional copyright. Copyright law is rooted in the notion of human creativity and authorship. Extending it to AI-generated material could muddy these foundational principles.

A sui generis system could be tailored specifically to the characteristics and challenges of AI-generated material. For instance, it could have a shorter term of protection than traditional copyright, reflecting the different nature and perhaps quicker obsolescence of AI-generated content. This separate system could also clearly define who holds the rights—the creator of the input or the user prompting the output. Much like a photograph, reality is a mix of both, and ultimately your rules must find the right balance to protect them both.

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

If so, how?

The Copyright Clause in the U.S. Constitution aims to "promote the progress of science and useful arts" by granting exclusive rights to authors and inventors. Permitting copyright protection for AI-generated material under the Copyright Clause is not necessary to encourage innovation in AI technology. AI advocates say that by providing legal protection and a potential revenue stream, developers might be incentivized to invest more resources into advancing AI systems, which could in turn promote scientific progress.

However, companies like OpenAI and Midjourney have already prove in the past year they can profit without legal copyright protections on the outputs.

It's important to note that the core purpose of copyright, as articulated in the Constitution, is to reward human creativity and innovation. Extending this protection to works generated by machines diverges from this original intent, and it’s important to only consider the people involved in each step of the process.

Generative AI model makers depend on the creators of their data the same way a book relies on its cover as an element. Both need to be on the same page as to how their collaborative work is licensed to end users. An output can’t be protected unless the entire chain of command it ethical. And please don’t use blockchain—it just adds more GPU environmental waste to the problem.

Infringement

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

All AI output is derivative of the data input into it. When a data model is made, they compress the data down to the latent space and then recall it back in mixes to create an output. It could take only trace amounts (0.001%) of 100,000 different works, but it’s still 100% derivative. This scale will be important because it helps visualize how each party’s rights should be balanced. Think of it like your collection rules because that’s what’s happening. AI companies aren’t stealing individual work—they’re stealing entire collections.

When I input the word “car” into an AI-image generator, the car would likely be trace amounts of 100,000 car images with maybe 10-12 images being above 1%.

When I type “car in the style of Pixar,” I’m heavily weighting Pixar work so that most (or possibly all) of those 100,000 images are Pixar’s. The image will have up to 10-15% similarity to images specifically for Pixar’s Cars film series.

When I train a LORA model on Lightning McQueen, I’m raising that percentage up over 20% or more to hyperfocus on that character.

It’s important that the original creator be properly compensated for each of these tiers, as there’s a distinct difference between how much of the original creator’s work I’m invoked in those hypothetical situations. This same concept applies to music, text, video, and code.

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

The substantial similarity test, a cornerstone in U.S. copyright law for determining infringement, may present challenges when applied to AI-generated outputs. In traditional cases, the test seeks to assess whether the defendant's work is sufficiently like the copyrighted work to conclude that unlawful appropriation has occurred. However, generative AI complicates this because it can mix elements from various sources, making it difficult to establish a one-to-one relationship of substantial similarity between the generated output and any single copyrighted work.

Instead, it's important to look at collections as described above. For example, if I make a video that's just a 5-second clip of every episode of The Simpsons, I'm only using a little big of each episode, but the entire video is still infringing on The Simpsons unless I go further to transform it with my own elements. AI is exactly this but on a mass scale—they did infringe when copying the data. That similarity needs to be zoomed into each digital work in the latent space so works are properly protected and everyone is properly compensated for their efforts.

24. How can copyright owners prove the element of copying (such as by demonstrating access to a copyrighted work) if the developer of the AI model does not maintain or make available records of what training material it used? Are existing civil discovery rules sufficient to address this situation?

Proving the element of copying in the context of AI-generated outputs can be extremely challenging if there's no record of the training materials used. In traditional copyright infringement cases, proof of access to the copyrighted material is often crucial to establish the likelihood of copying. Without access to records of what training materials were used, copyright owners face a significant hurdle.

Existing civil discovery rules might be insufficient for tackling this situation effectively. The process of discovery in AI-related cases could quickly become complicated and burdensome for both parties. The sheer volume and diversity of data used for training, the complexity of AI models, and the proprietary nature of many of these models all contribute to the problem. Existing discovery rules are not inherently designed to handle these challenges and may not provide the level of granularity needed to trace back a particular output to a specific copyrighted work used in training.

This lack of clarity and the limitations of existing discovery rules could either result in undue burdens on AI developers to prove non-infringement or make it virtually impossible for copyright owners to prove infringement. In both scenarios, the current legal framework appears inadequate for the nuanced challenges presented by AI technologies.

To address these gaps, there may be a need for specialized guidelines or modifications to current discovery rules to facilitate more effective, efficient, and fair processes in cases involving AI and copyright. Whether that takes the form of legislation, amendments to procedural rules, or development of new judicial precedents is a matter for further discussion and consideration. Think of automobiles—removing the VIN from a car is in itself a crime with penalties. Not disclosing the datasets should be considered a form of removing CMI, which is infringing.

Much like the IRS requires you save tax information for three years, it's important that AI companies maintain the CMI for their data so they can prove if it is properly licensed or not.

25. If AI-generated material is found to infringe a copyrighted work, who should be directly or secondarily liable—the developer of a generative AI model, the developer of the system incorporating that model, end users of the system, or other parties?

Determining liability in the case of AI-generated material found to infringe a copyrighted work is a complex issue that can vary depending on the circumstances. Here are some potential scenarios:

Developer of the Generative AI Model: If it can be proven that the model was specifically designed to replicate or mimic copyrighted material without consent/compensation/credit, then the developer of that model should be held directly liable. In addition, if it's proven that the model contains references to copyrighted work at all used without consent/compensation/credit (see HaveIBeenTrained) then they should also be held liable. This would cover infringing works in LAION datasets.

Developer of the System Incorporating that Model: If this entity customizes or fine-tunes the model in such a way that it produces infringing material, they could also be directly liable. In other cases, they might face secondary liability if they were aware that the AI model they incorporated had a high likelihood of generating infringing material but used it nonetheless. This includes MidJourney, Stability Ai, OpenAI, Anthropic, and others using these models for commercial purposes by selling subscriptions.

End Users: If end users utilize the system to deliberately generate infringing material, they could be held directly liable for their actions. This would likely depend on their knowledge and intent, as evidenced by using names of artists or their protected works. It should also cover infringing models on Civit AI.

Other Parties: Depending on the situation, other parties like distributors or platforms that host AI-generated material could also be implicated under theories of secondary liability if they had the right and ability to control the infringing activity and had a financial interest in it. This includes platforms like Hugging Face and GitHub.

To achieve clarity on this issue, legislative or judicial guidelines is needed. As AI technologies become increasingly integrated into various aspects of life and business, the importance of clear rules allocating responsibility and liability will only grow. Without them, the risks associated with developing, deploying, or using generative AI systems could become a deterrent to innovation and economic growth.

25.1. Do “open-source” AI models raise unique considerations with respect to infringement based on their outputs?

Open-source AI models do present unique considerations with respect to infringement. The collaborative and decentralized nature of open-source development makes it more difficult to pinpoint liability. Open-source projects often have many contributors, and it can be difficult to ascertain who is responsible for any given piece of code that might lead to infringement.

Since open-source models are publicly available, anyone can modify them. A user could make modifications that lead to infringing outputs, yet the original developers may be unaware of these actions.

Developers who release their models as open source often relinquish control over how the models are used. This could complicate the issue of secondary liability because one could argue they lack the "right and ability to control" infringing activity. The communal nature of open source also dilutes the level of due diligence exercised in ensuring that the model does not infringe upon existing copyrights. With proprietary models, a single entity is generally responsible for vetting content and can be held accountable.

Because of these factors, assigning liability for copyright infringement in the context of open-source AI models is particularly challenging. Legal frameworks may need to adapt to these unique circumstances.

26. If a generative AI system is trained on copyrighted works containing copyright management information, how does 17 U.S.C. 1202(b) apply to the treatment of that information in outputs of the system?

17 U.S.C. § 1202(b) makes it illegal to "intentionally remove or alter any copyright management information," or to distribute works knowing that copyright management information has been removed or altered without proper authority. The statute defines copyright management information to include information like the title of the work, the author, and terms for use of the work, among other things.

If AI companies are not fully transparent about what data is contained within their datasets, then they are guilty of removing this information, regardless of whether they ban the words being used by end users. This is the exact clause that should be used to enforce transparency, as not providing it means they are violating this law. Once provided, it should be easy enough to prove whether copyrighted works were used in training, thus violating copyright law.

27. Please describe any other issues that you believe policymakers should consider with respect to potential copyright liability based on AI-generated output.

Policymakers should consider the following issues related to potential copyright liability based on AI-generated output:

Algorithmic Bias: If an AI system is trained on copyrighted works that reflect certain viewpoints, cultures, or styles, there's a risk that its outputs could perpetuate those biases. This raises questions about who is responsible for any ensuing copyright violations or ethical concerns. I believe the model makers, the companies training algorithms on them, and those implementing these algorithms need to be held accountable.

Shared Liability: Multiple parties may be involved in training, developing, deploying, and using AI systems. It's vital that you lay out the exact responsibilities at each end.

Ethical Use: Beyond just the question of legal liability, there's also the issue of ethical obligations — both on the part of the people training the AI systems and those using them. Applications such as the Undress App or Mr Deepfakes are not ethical and need to be addressed as well. Even companies like Adobe, Shutterstock, and Getty that claim to be doing things ethically are not following the 4 pillars.

Copyright Trolls: Like other areas of copyright law, AI-generated content is also susceptible to actions from copyright trolls. This could be an unintended consequence of allowing copyright protections on AI-generated outputs, and it's important that we don't allow AI companies to destroy the original intent of copyright laws, which is to protect our creative works to encourage meaningful science and art.

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 law should indeed require AI-generated material to be labeled or publicly identified as such. This kind of transparency would serve multiple purposes, including informing consumers, protecting copyright holders, and guiding legal and ethical oversight.

In practical terms, a digital watermark or metadata could be embedded within the AI-generated content, or an accompanying label could be displayed whenever the content is shown. This requirement could be especially crucial in contexts where the distinction between human-created and AI-generated content has significant implications, such as in news journalism, academic research, or legal documents.

Consumers should have the right to choose human or machine for our content. By clearly labeling AI-generated material, we can establish a more trustworthy environment for both creators and consumers, enhancing the credibility and ethical standing of AI systems.

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

The responsibility for identifying a work as AI-generated should primarily fall on the entity that deploys the AI system to generate the content. Whether it's a company, a research institution, or an individual developer, the onus should be on them to ensure that the AI-generated nature of the work is clearly labeled.

In cases where multiple parties are involved—such as a dataset curator, an AI model developer, and a commercial user of the model—each party, including the end user who publishes or disseminates the AI-generated content, should bear the responsibility for labeling it appropriately. This would ensure that the end consumer or viewer is well-informed about the nature of the content they are interacting with.

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

Yes, there are barriers to labeling or identification requirements for AI-generated material. One technical challenge is ensuring that the label remains intact and is not easily removable, especially when the content undergoes further modifications or is shared across various platforms. Metadata can be stripped or altered, and watermarks can sometimes be removed.

From a practical standpoint, the issue becomes more complex when considering the different types of content that AI can generate, ranging from text and images to videos and even code. Each type of content may require a different method for labeling, and it would need to be both noticeable yet unobtrusive. Moreover, in a collaborative environment where both human and AI-generated content are combined, determining what percentage of AI contribution necessitates a label could become a contentious issue.

Despite these challenges, the overarching goal should be transparency for the end-user. Even with these barriers, efforts should be made to develop robust methods for labeling AI-generated content in a way that balances practicality and the public's right to know. Even if there are specific labels like “made by AI” versus “made with AI” at certain thresholds, anything will help so long as we all understand how to follow and enforce the rules.

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?

Failure to label an AI-generated work, or the intentional removal of such a label, could undermine transparency and mislead consumers or users. Serious and consequential legal penalties must be imposed for such omissions or actions. These penalties should range from fines to more serious sanctions, depending on the severity and intent behind the non-compliance. The consequences could be more severe in cases where the lack of labeling leads to significant harm or deception.

The nature of the consequences would likely need to be tailored to the type of work and the context in which it is used. For instance, failure to label AI-generated content that is part of a news article may be considered more serious than omitting a label from a piece of AI-generated art. However, both should be considered serious violations of the law.

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?

Several tools and methods are under development or already in use for identifying AI-generated material. These include:

Deep Learning Classifiers: These machine learning models are trained to distinguish between human-created and AI-generated content, though they should not be used due to their strong biases, such as against non-native English writers. Both image and writing detection algortihms have proven flaws and should not be trusted.

Digital Watermarking and Metadata: Some AI models can include a form of digital watermark in the generated content to signify its AI origin. Traditional digital forensic techniques can sometimes be adapted to identify the tell-tale signs of AI-generated content, such as inconsistencies in lighting or texture that are hard for current AI models to get right. Although watermarks and can be removed or altered, they're as important as any other CMI and can be easily added to that section of the existing law.

Forensic Methods: Traditional digital forensic techniques can sometimes be adapted to identify the tell-tale signs of AI-generated content, such as inconsistencies in lighting or texture that are hard for current AI models to get right. Unique signatures in the material can also indicate whether it's AI-generated. For example, Deepfake detection tools often look for inconsistencies in blinking or breathing patterns in videos.

Crowdsourced Verification: Open platforms where users flag AI-generated content can also serve as a tool, though they can be prone to errors and abuse. I am more than happy to volunteer our services at Luddite Pro, along with all the other advocacy groups who have actively been working on social platforms like Twitter, Reddit, Instagram, Pinterest, and Facebook to identify and call out AI-generated content that's not properly disclosed.

Additional Questions About Issues Related to Copyright

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?

The legal rights that currently apply to AI-generated material featuring the name or likeness, including vocal likeness, of a particular person generally fall under the categories of right of publicity and privacy laws, defamation, and sometimes even false light claims. These rights can vary significantly by jurisdiction, both within the United States and internationally.

Right of Publicity: We could use a federal right of publicity, which gives people the right to control the commercial use of their name, image, and likeness. Unauthorized use of a person's likeness in an AI-generated work could therefore be subject to legal action. Right now, it's only available in certain states.

Privacy Laws: Privacy laws protect individuals against the unauthorized dissemination of their private information. These laws should be invoked if AI-generated material were to include personal details without consent.

Defamation: If AI-generated material falsely portrays a person in a way that damages their reputation, there should be grounds for a defamation suit.

Trademarks: If an individual’s name or likeness has been trademarked, then unauthorized use could constitute trademark infringement. This protection should also apply to copyrights, as well as individual rights to publicity.

Other Federal Laws: In some circumstances, federal laws like the Lanham Act, which governs trademarks and unfair competition, could also come into play, especially if the AI-generated material is used in a way that causes consumer confusion.

These are complex legal areas with many nuances, and they're often subject to ongoing litigation and legislative changes. The intersection of these laws with AI-generated material is an emerging area of legal concern and is likely to see significant developments in the coming years.

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?

If Congress decides to establish a new federal right that's similar to state law rights of publicity, it would offer a standardized legal framework across the United States. This standardization would be particularly useful given the complexities associated with AI technologies that span multiple jurisdictions. Setting a minimum standard allows individual states to add further protections if they see fit, which may be the most beneficial setup for consumers and copyright holders alike.

The scope of this federal right could extend to cover the commercial use of an individual’s name, image, and vocal likeness, regardless of whether it’s used in AI-generated materials. This would require outlining the specific conditions under which consent is needed and enumerating exceptions such as uses for journalism or academic research. This already exists in existing copyright law and would go a long way towards protecting our personal individuality, regardless of whether we choose to commercialize it.

Enforcement mechanisms would also need to be specified, through both civil remedies and criminal penalties. Given the rapid speed at which AI-generated material can be created and distributed, it is crucial for enforcement measures to be timely and effective. Ethical concerns like the rise of deepfakes and other types of manipulated media that could deceive or harm individuals or the public should also be considered in the legislation.

Special considerations might also be necessary for deceased individuals, minors, and public figures, as each category presents its own unique issues. Implementing a federal right would require an extensive discussion involving legal experts, ethicists, technologists, and the general public to ensure an appropriate balance between fostering innovation and protecting individual rights. Please contact everyone mentioned above.

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—although it’s not an issue of copyrighting an artist’s style. For example, the artistic styles of both Pixar and Dreamworks are very similar. If I happen to make something in a similar style, it should not be infringing. However, if I use the prompt “in the style of Pixar and Dreamworks” or I train a model entirely on Pixar or Dreamworks content, then I am obviously infringing based on the weights illustrated above.

There is a quantifiable difference between how much of an artist's work I'm infringing upon when I say "mouse' versus "Mickey Mouse' and "mouse in the style of Disney." In two of those cases, I am directly infringing by using protected IP as a shortcut. That shortcut should be protected, and this would apply evenly to images in the style of an artist, chatbots in the style of any person/character (see Character AI), voices, music, movies, and more.

The problem isn't making something in the style of (person/brand). The problem is making it using the name of (person/brand). It can also come up if you feed Disney work into your model to train it. The style of Disney isn't what needs to be protected—the protection occurs in the training data and in limiting the use of other people's names/work in prompts or image-to-image training.

In any case, careful consideration would be needed to balance the interests of human creators, who deserve recognition and reward for their unique contributions, against broader societal interests in fostering artistic and technological innovation. It would likely necessitate a multi-disciplinary approach, involving not just lawmakers but also artists, technologists, and ethicists, to formulate a legal framework that is both fair and practical.

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?

Section 114(b) of the Copyright Act governs the limitations on exclusive rights related to sound recordings. It defines the scope of what is protected under a sound recording's copyright and what isn't, essentially delineating the borders between federal copyright law and state laws, including right of publicity laws. State right of publicity laws can protect the commercial use of a person's name, likeness, or voice, but they often intersect and sometimes clash with federal copyright protections.

In the context of generative AI, which can generate sound recordings that mimic the voice or style of a particular artist or personality, it lives at the intersection between Section 114(b) and state rights of publicity. If an AI-generated sound recording mimics the voice of a particular artist, it infringes upon the artist's rights, following the precedent set by Midler vs Ford Motor Co.

34. Please identify any issues not mentioned above that the Copyright Office should consider in conducting this study.

Do your best to strike the balance between providing fair compensation/credit/consent for existing IP holders, along with reiterating our abilities to receive copyright protection immediately upon creating our work. Be sure to clearly discern between the levels of participation between all parties at each level.

If you do this, you should be on the right track and set a precedent that will have positive outcomes for all involved.