

I will answer each question posed for this survey:

1. The value of art copyrightable by a human author

AI tools are a new way that original art can be created in essentially all artistic mediums. They will fundamentally impact all industries they touch. Just like 3D rendering software changed numerous industries; for the worse in many cases of prior existing industries. As for value, it should be self-evident if you use any of the common tools like Midjourney. There are 2 million career artists in the US. Worldwide there are estimates in the range of 10-20 million people. I'm sure many more people have dabbled or do art on the side. Just Midjourney has nearly 15 million users at the time of writing. There are many other services and technologies that have similar numbers of users. ChatGPT has 100 million users. These are tools that millions of people are using to develop their own art. Despite what many bigots would want you to think, large numbers of these users are making art that no one else will make, that is unique to them and their life experience, where they are able to express themselves creatively in ways they never could have otherwise.

2. Does AI raise uniqueness issues for existing copyright holders?

Of course, you will find examples of this in *any* creative industry that gets the bar to entry lowered. I do illustrations, I do programming, I do 3D modeling. All of these industries will have "uniqueness" issues, and historically have had uniqueness issues with any new tool introduced. This is a poor question. The creation of cheap game engines like Unity3D and Unreal Engine historically created massive uniqueness issues across the entire gaming sector as well as 3D animation and other tangential sectors. However, they also made the tools available to creatives who otherwise would have never been able to realize their ambitions. Now many people play those 'indie' titles to great enjoyment and the developers of those are compensated well if they are able to make a mark on the market.

3. Relevant papers and studies

<https://copyright.gov/comp3/chap900/ch900-visual-art.pdf> ch 909.1

Lays out why AI art should be copyrightable so long as there is creativity in the creation of the work. Also lays out that it doesn't matter if other people can reproduce similar works, everyone gets copyright so long as they each had creative control over their shot. Lays out specific requirements that many AI art pieces meet and surpass. When you are rendering an image with stable diffusion, you specify subject matter, lighting, positioning of the subject matter, you choose the model to use which influences the final style, you make many other creative decisions. If photography has such low limits to creative matters that I can go walk down the street, take a picture on my smartphone, and get copyright; there is no argument grounded in

reality that can justify how an artist using AI tools with very specific descriptions of many pieces of a scene and rendering that cannot get copyright.

4. Relevant legislation

<https://www.reuters.com/technology/japan-leaning-toward-softer-ai-rules-than-eu-source-2023-07-03/>

Why is the copyright office asking random people on the internet to do due diligence for it? You should know this without thousands of people who don't know anything spamming you.

5. Is new legislation needed?

Apparently, considering the copyright office doesn't follow its own rules. There is no question of copyright when I change some numbers in Blender -- some 3D rendering software, and render an image of a few boxes lit by a light. There is no question of copyright when I walk down a river and take a few dozen pictures of it. Yet for some reason the copyright office has denied copyright to works that have much more creative input and iteration than either of these. This should be explicitly spelled out that AI art is no different. Because apparently, for some reason, this is not self-evident to a lot of people who want to disparage artists' hard work.

Zarya of the Dawn should be given copyright over the images in the graphic novel as a start. The copyright office clearly failed to understand how the tools were used. And it set a precedent that gave online trolls reason to harass people making art with AI tools, and many artists who never used AI tools to make their art, whose art happened to look kind of like some AI art.

It should be stated explicitly that any use of tools, no matter how "creative" they seem, is ultimately a use of tools. The UK has had such legislation for many decades

<https://www.legislation.gov.uk/ukpga/1988/48/section/9>

"In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken."

6. What kinds of copyright-protected training materials are used

Any public data on any copyrighted thing you could imagine. These materials are typically collected by web scraping public website data and curated automatically using machine learning algorithms that categorize and rate the content.

6.1 How or where do developers collect their dataset data?

Same as above, they scrape it in cases that need a lot of data. Researchers and more specific things make their own data due to a lack of public data on specific topics. For instance, a bank may want to detect fraud on accounts they control. They may collect instances where they found an account to be compromised and label that as something they want to detect, and then feed that along with many other accounts' activity that was not fraudulent to teach it to classify if activity is fraudulent or not.

6.2 What licenses are used for current copyrighted training data

I'm sure these are few and far between things for *private* training data. Some websites have said their content cannot be used for training AI. This is tantamount to asking Google to not index their websites since Google has operated its search engine using machine learning for 2.5 decades at this point. If they cannot train on the content provided, they cannot index the content.

6.3 Extent of non-copyrighted data usage

It's not possible to tell realistically how much public domain data is used, and how much licensed data is used. I can tell you that Adobe Firefly is entirely trained on data that is either public domain, or they have licensed for Adobe Stock. And it is competitive with any stable diffusion model trained on publicly posted images. There is a route you could say that things need a license to be used and can't be scraped, but that would be against the public's best interests since it would just enforce monopolies by Adobe and other companies on tools. Meanwhile Adobe will be impacting the market regardless of if there are open source alternatives.

6.4 Data retention

It depends on the developer for their data retention. I will reiterate, this is public data primarily. Asking about data retention is like asking if anyone viewing a site could've saved a picture on it to their phone. There isn't value in the discussion. In the first place many developers never actually have the data other than during the moment it's going through the training process. Some of the biggest data sets are only links to the source content.

7.1 How are training materials used/reproduced?

There have been multiple studies that have demonstrated that models are capable of reproducing works if you prompt them enough with the exact input that describes an image that is repeated thousands of times in the source dataset. And then you proceed to attempt to copy *something* a few million times. For example <https://arxiv.org/pdf/2301.13188.pdf>

This paper shows how if you are trying to use the model to copy a work, after a few million tries it may create something that looks vaguely similar to one of the input images.

I can happen to reproduce an image I've never seen before quite closely if you give me a few million tries.

This question is weird. Just because something is technically capable of reproducing something doesn't make it a copyright issue. If this were an actual valid question, it'd be an issue that models could reproduce copyrighted characters like Spiderman. Or that I am technically capable of violating copyright by producing a drawing of Spiderman.

An important note here is that it is trivial to make models generate reproductions of specific images that never existed in their training set. If you give them a description that's specific enough they will reproduce the image. These are referred to as textual embeddings.

This is all to say, it doesn't matter if they can, given infinite monkeys with typewriters, reproduce the copyrighted works of some artists. What matters is the copyright infringement done by the user.

7.2 How are inferences stored or represented in the AI model

This is a large topic, [like the biological brains these models are modeled after](#) they store inferences in various ways. They are capable of memorizing exact things if you try to make them do that, but as shown above, if you're not trying, you're not going to have them memorize things.

The important thing here is that, like human brains, they will find common patterns and associate them. If you show an AI one picture of a dog and tell it's a dog over and over, it will think that one picture is a dog and if you show it a different picture of a different dog breed it might not understand that it is also a dog. You can do the same with a human child. However if you show it a lot of pictures of dogs and tell it they are all dogs, it will gain a semantic understanding of what a dog is. Just like a human.

If you show me a lot of pictures of some specific artist's art, I will start to be able to pick out what features that particular artist uses frequently. How they use color, how they tend to compose their works, what kind of brushes or medium they use for their art, how they depict human features, etc. That singular name gains a large semantic meaning to me that encompasses a lot of information about that artist. AI is the same. You can tell this if you either have any understanding of how machine learning works, or have used the existing tools like midjourney much at all.

7.3 Is it possible for a model to “unlearn” things it learned

Yes, sort of, but also that's not how art or language works. There was an interesting paper recently titled “Erasing Concepts from Diffusion Models” <https://arxiv.org/pdf/2303.07345.pdf>

In this paper they demonstrate the capability to erase concepts like art styles, as well as the concept of objects like cars. Which is very cool, except that they demonstrate two key points.

1. They are primarily erasing the language association

This means that while maybe I can't reproduce Starry Night using the term, Starry Night, I may be able to reproduce it by describing it more specifically. And existing models have tools to specifically describe images in a way models can interpret.



Figure C.1: Erasing a single artwork image does not effect the corresponding artist style or other memorized artworks. The edited model after erasing "Starry Night" clearly has minimal effect on Van Gogh style and other memorized artwork while effectively erasing "Starry Night".

2. They have some bleed where erasing one concept can negatively impact other similar styles



This actually just makes it infeasible in practice. This would essentially mean that erasing an artist from a network erases the ability to do any art that looks similar to theirs in many cases. Effectively saying for example, that a particular watercolor artist gets ownership over the style of “watercolored paintings”

7.4 If you don't have access to a training set is it possible to identify if an artist's work is in the training set?

Sort of, not really in the way you'd be able to verify how meaningfully they are in the training set.

For example, Stability AI at the end of 2022 released Stable Diffusion 2.0. When Stability was making Stable Diffusion 2.0, they found that removing the data for an artist who had gained popularity as a description for his art style, Greg Rutkowski, along with many other artists, did not actually stop Stable Diffusion from being able to imitate their styles.

This is where the two main pieces of a diffusion network come into play. One piece renders the image from noise, this is where it gets its diffusion name. However, that model on its own doesn't really know how to translate a sentence into what the sentence actually means. Like I can say "a picture of a bright blue apple" and it is unlikely the model has much training data on blue apples and it probably doesn't understand the relationship between the color blue and the apple much. This is where the other main network comes in, the textual embedding network. In the original Stable Diffusion this was [CLIP](#) trained by OpenAI. The only job of this network is to reduce a few words into a set of a few thousand numbers that encompass the meaning of the sentence.

Now it may be clear why the Stable Diffusion model was still rendering works in similar styles to these artists. The artists were not represented in the image training set for actually rendering images. They were represented in the data set used to train the thing that only describes images. Simply, their names had become synonymous with saying something like "painterly art style with high contrast details, dragons, magic, etc" and that was all it took for Stable Diffusion to reproduce their artworks.

Stability then trained an alternative called OpenCLIP without their names in the dataset and they were able to avoid the artist being used for a few days until someone trained a LoRA that could describe the style easily again.

So yes, you can in some cases tell if an artist's work was in a training set if you can get their names to invoke enough of a response to assume as such. That is not hard proof though. And it's dubious if the model just knows their name describes a more rich meaning.

8. What cases would the unauthorized use of copyrighted works constitute fair use

Fair use, among its many factors, gives heavy weight to how transformative the use of the work is. The use of materials in training is so transformative that it isn't functionally understood to be even considered by fair use. The individual contribution of an artwork to training amounts to someone seeing a picture of an apple and reaffirming that it is indeed an apple. It is even further

down the scale than people specifically learning a style or how to draw a specific object. A person would look at a handful of reference images for an object, AI training would look at tens of thousands. Using current reasoning on fair use paints a very clear picture that if AI models are not to be transformative enough for fair use, then humans are not transformative enough to make original art.

There is some case for impact on the particular artist's business that depends on their style, but you could make those arguments even if that artist wasn't in the training set. That is to say, a particular artist's copyrighted works are immaterial in their ability to be used to impact any given artist. Copyright doesn't cover things like copyrighting a style for good reason.

I will link cases I'm sure you're aware of, but are failing to note intentionally:

https://en.wikipedia.org/wiki/Authors_Guild,_Inc._v._Google,_Inc.

https://en.wikipedia.org/wiki/Perfect_10,_Inc._v._Amazon.com,_Inc.

8.1 Purpose and character

The purpose is to create tools that make artists' work easier. It is making a piece of software that allows users to create works that are similar, or completely different. Many of these models and software are free and open source.

AI tools are objectively a massive benefit creatively to both new artists and existing artists. They are the modern day digital art software to the traditional mediums of art.

It's possible that finetunes of particular copyrighted characters could be found to not be fair use. They are training to describe a copyrighted character after all. However on the other hand if you were to make a finetune on the art style of some particular artist. First art styles aren't copyrightable so unless you are using the model to directly reproduce the exact artist's works there isn't a copyright-centric way of saying that's not allowed. I will reiterate the earlier question of "can you tell if an artist's works were used to train a model" in a different way and ask, if a finetune says "style of painting and dragons" could you determine that it's using Greg Rutkowski's works? The answer is no you could not. Copyright is not made to control these kinds of things, it cannot be under how it works, and it should not be as that means anyone could sue an artist for imitating their style, claiming an AI model was trained on their style. The defendant would have no way to prove reasonably that they did not do that at any point.

8.2 How should this apply to distributors of copyrighted content that's used in training, that the distributors do not use themselves?

Well first of all, that means that basically any online content publishing service that publishes user content publicly could be found liable if you were to argue that both it's copyright

infringement and that it could apply to distributors. Reddit, which has been scraped by many models for its content? Oh well Reddit shouldn't have published that user content.

But I digress, the point I made above is that it doesn't make sense to have copyright apply outside of the bounds of copyright so this question is meaningless in that framework.

8.3 What happens if a model is adopted for commercial purposes?

Nothing. The data is public. The model is transformative beyond all measure. The individual contribution of data to a model is so vanishingly small that it's only possible in very rare cases to have a person's art influence the output in a meaningful way if they are both very famous, and if the end user invokes their identity specifically.

Again, Authors Guild v. Google demonstrated that even though Google was making money off of indexing books for Google Books, the value provided and transformative nature of the use outweighed any financial interests they had.

8.4 What quantity do model trainers use for training, does it affect the fair use analysis?

Models typically use vast quantities of data. Stable Diffusion was trained on 5 Billion images. Due to the nature of the models, if the sample size for the data dropped anywhere close to the magnitudes where it might bring the transformative nature into question for large models trained by big companies, the models likely wouldn't be useful.

I will pose a different question. If a team of artists is told to imitate the art style of some famous artist, but that artist is not paid what happens? Yes that's right it's fine because the artist cannot copyright their style. That team of artists can be looking at 10's or even a couple images that encapsulate that artist's art style. But any art those artists generate is not bound by the copyright of that artist they are imitating the style of.

This is not uncommon to have artists told to imitate something specifically. The movie industry does this frequently, with temp tracks that are a specific song they don't have license to, but want to imitate as closely as possible. They often have different artists make a song very similar to that single song for the movie.

The various animation industries do this as well. Asking large teams of artists to specifically imitate some style of a given artist.

Fundamentally the process of training a model is incredibly transformative no matter how little content there is. Just like a human looking at a picture, understanding what you mean when you say "I want that style to be replicated."

If we look at this from the impact on the market for the artist and were to assume you couldn't make a model describe a style, it's the same if you use a model or a human, you're just paying the human some money to launder the style.

8.5 Fair use effect on potential market

This should be looked at as a question of what copyright is intended to protect. 3 options are posed by this question "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? "

All of these are questions to ask. And all of them become meaningless unless you wish to copyright styles. AI models are a tool. They are a tool that will exist and be capable of imitating any author's style or works if given the proper description. They do not need any particular artist's works to emulate their style. They can make new styles that are interpolations of the various things that encompass a "style."

So if the question is "do the models compete with the copyrighted work" then the answer is, yes they can, so can copy and paste on your computer.

If the question is "do the models compete with the body of works of an author" the answer is, yes, they can, regardless of if the particular author's works were used to train them at any step.

If the question is "do they compete with the general class of works is", you guessed it, yes, they can, any new tool that brings the bar to entry down will do that.

So what would enforcing copyright at any point in this hierarchy serve to do? It wouldn't protect artists because their styles are trivially reproducible by models even if you don't train on their works. It wouldn't make sense to protect their industry since the tools will still disrupt it even if they must only be trained on public domain works.

Again, Adobe Firefly is trained on public domain works and works that Adobe licensed. It will disrupt all of these markets without laying a finger on any artist's copyrighted works.

So again, what will this protect? Not the artist that's for sure, an artist can't prove their works were used to train a model. Someone imitating their style could've just described their style closely enough. What it does protect is monopolies. Now only companies that can license large volumes of work would be able to make the tools. It would also stifle innovation and cause a chilling effect on art as a whole. Now you can't post art without being open to accusations of stealing a style. Even if you're not using AI tools, the people making accusations don't know that.

9. Should copyright owners opt-out or opt-in to model training

If a thing is posted in public the copyright owners give implicit license for people to look at their work. The platform they post on typically has a license to redistribute their work. The act of posting their work in public means it may be used for training. That's just how the internet works. If you don't want people to look at your works and learn from them, don't post them. Keep them to yourself or in private distribution channels that are not public.

9.1 Should consent be required in commercial use cases

N/A, consent is given by posting the work publically

9.2 How would an opt out system work?

It wouldn't. Neither would an opt-in system. The data needs to not be in the training set to start with. The moment someone reposts a public work, any and all tracking of opt in or opt out is gone. Any slight transformation to the image data breaks databases of "opt-in" or "opt-out" content as well. Opt-in/out fundamentally doesn't make sense from a copyright or a technology standpoint.

There isn't opt-in or opt-out to fair use.

9.3 Are there obstacles to making such a system work?

Yes, as noted above. Companies would just opt to make their own datasets like Adobe because the alternatives don't make sense. Or they would use works from countries that don't have such asinine copyright laws.

9.4 What happens if an objection is not honored

N/A

If there is legitimate private information in the training set that got into it because someone accidentally leaked that personal information, that should be the only reasonable time that it can be excluded from the data set. And that is outside of the scope of the copyright office.

9.5 What happens with work made for hire or otherwise work not owned by the creator

No. This is a worrying question or line of thought. I don't know why the copyright office would even suggest that someone has some modicum of copyright on something they explicitly signed

copyright away on. Work made for hire is considered authored and owned by the one doing the hiring. This is an insane proposition. If an author does not want their works used in a thing they would be signing away copyright to, do not sign away copyright in the first place.

I'm just dumbfounded how this is even a question.

10. If consent is required, how is the license handled?

N/A

If you want advice on licenses ask the companies doing them now like Adobe and OpenAI. A license should not be required to train public content. For private content companies have been doing this forever.

11. What issues would there be with obtaining licenses for training?

There probably wouldn't be big issues for big companies, but their licensing terms would likely pay out on the order of dollars per work.

12. Is it possible to determine how much a particular work has contributed to a model

No, you can say the proof is in the pudding and the model being capable of being influenced towards a work by prompting for it counts as contribution. But as stated before you can't tell. Maybe your work that's a picture of a hat in some nice lighting next to the beach contributed. But it just contributed to the semantic understanding of what a hat is, along with 10 million other images.

13. What's the economical impact of licensing for developing AI systems

Authors would get shafted since their work isn't worth much to a company with hundreds of thousands of other people who would gladly license their library of works to train on for a few hundred dollars flat. Or the companies would pay out based on model use, but the contributions would be so minimal that payout would be dictated by how many works you submit and it'd still amount to a few hundred dollars over your lifetime unless you submit hundreds of thousands of works.

Monopolies would build their empires more on being able to monopolize AI systems for any professional work.

Due to the companies having a monopoly on the tools, they have a monopoly on the data they get from users using the tools. So they can train a positive feedback loop using [RLHF](#) and eventually stop needing to pay out as many artists since they can create better models with less data. Companies are already doing this only on the basis that it's costly to train in the first place. If it becomes an insurmountable barrier due to licensing fees, it will be impossible to compete with.

Creators who want to use AI tools to empower their work and compete with others would be forced to buy into tools with likely inflated prices due to the monopoly.

This would actively damage all creative fields and only serve to channel money to companies like Adobe and OpenAI.

14. Any additional factors

Not that I am thinking of at the moment :)

15. Should trainers of models keep records of their datasets?

Anyone that's training large models likely already does. LAION is just a list of links to the original content.

It'd be nice to see more public info about data used to train models for the sake of competition, though as time goes on that data will skew into private data used to train with systems like RLHF.

16. What obligations are there to notify copyright holders that their works have been used to train a model

Should copyright holders be notified if I link the image they posted publicly to Reddit to my friends? That would just mean they get spammed with hundreds of thousands of notifications wouldn't it? That just seems bad for the copyright holder to be honest.

17. Are there any existing U.S. laws outside of copyright that could require developers to disclose records about what they used for training?

Maybe? Not that I know of personally. This question really makes it sound like the copyright office is reaching to try and meet a certain conclusion though.

18. Under copyright law are there circumstances where a human using generative AI would be considered the author

[https://copyright.gov/comp3/chap900/ch900-visual-art.pdf ch 909.1](https://copyright.gov/comp3/chap900/ch900-visual-art.pdf)

Yes. As stated above the copyright office's own guidance contradicts rulings they have provided. If AI art cannot be copyrighted without major creative input then you should consider increasing the bar for a modicum of creativity in photography.

Let's ask some questions here about what constitutes creative work. Because it seems like the office is not clear on what's established.

Let's say I open up the Stable Diffusion interface I use. It appears that the office believes there is no, or essentially no creative input. So what happens if I have essentially no creative input? So I open my stable diffusion interface and leave it on its defaults. I use the stable diffusion 1.4 model, and generate 9 images. And I get...



Oh, that doesn't look very interesting... That's certainly not anything I'd want to show anyone. That's what I get when I exert 0 creative input. Which seems to be what the copyright office believes people using these tools do.

I personally like rivers a lot, especially rivers at sunset. I frequently walk by a river near my house and take pictures of it that would be copyrightable. Some might say that my life experiences influence my like of rivers.

I also like anime, it'd be cool if I could render some images of a river at sunset with an illustration style. Let's try that!

So I choose an anime model called CetusMix and I will try making some pictures of a river.

After some iteration I end up prompting it with "Wide shot of a river at sunset, bloom, water, river, (wide shot:1.3), night sky, starry sky, grass, rock, cloud, sunset, dusk, scenery, trees, perspective"

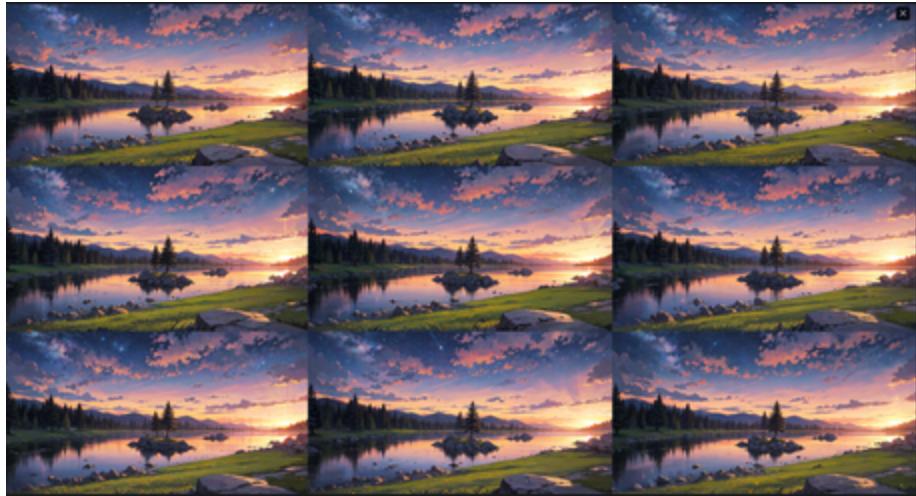
I capture a few dozen pictures, and finally I spot one from this set that I like



Specifically I like the one in the middle.



I then iterate on this creating many higher resolution versions of it, iteratively increasing the resolution and picking the ones that I like most while sometimes inpainting out details that I don't like or want to change.



I happened to get one where a shooting star was falling, that's great. But at the same time one of the trees moved into the background and I don't like that.



So I will edit that out using inpainting.



There we go, that's pretty nice in my opinion. I would go on to do more iterations of higher resolutions, usually ending in my opening my favorite image software Clip Studio Paint to do some color adjustments before I'm done. For stuff that I typically make, I literally render hundreds to thousands of iterations before I make something I'm happy with posting publicly.

But alas, according to the copyright office's ruling on Zarya of the Dawn. This would not qualify for copyright and despite my specification of framing, subject matter, and lighting, my selection process, my selection of model, and my selection of style. All things that would qualify a photograph easily multiple times over.

Well of course it can't be creative because it's random right??? But it's not random... At least not more random than the pictures I take on my phone where I don't control the placement of trees, where I don't control the placement of my camera along a river much, where I don't control the lighting and am subject to the weather. Remember, we know what random looks like. It looks like this:



As the copyright office you cannot justify this on the basis you have used in the past. You must justify this on some other reasoning other than creativity because your argument will fall apart the moment this goes in front of a half competent judge in court.

19. Are there any revisions to the copyright act?

I will again link <https://www.legislation.gov.uk/ukpga/1988/48/section/9> which should be part of US copyright law since it seems like the office can't recognize that using computer tools as part of the creative process only works when there's a human doing creative things.

As recently upheld correctly, an AI cannot have copyright on something, just as you cannot copyright something that God has written.

<https://www.documentcloud.org/documents/23919666-thalervperlmutter?responsive=1&title=1>

You need human authorship to have copyright. And you have human authorship when using AI tools unless you are literally attempting to have no creative input. These are works that I had creative intent to produce. It does not matter if they match an exact layout, photographs, or even lines I draw don't follow what I want exactly, and that's okay.

20. Is legal protection for AI-generated material desirable?

Yes, if AI-generated material is not protected it means you now have plausible deniability for any copyright infringement in existence. "Oh we thought that was AI-generated work and okay to use."

If it's protected then that gives companies massive incentive to improve tools and allow creators to have much more capability to create content. Creators also are incentivized to create high quality content.

20.1 Should protection be part of copyright or its own thing

Why would it not be part of copyright... This is all stuff that should obviously be covered under copyright and if it isn't then current measures for copyright should be reevaluated for *all* copyrightable works.

21. Does the copyright clause of the constitution permit AI-generated material?

What? Yes? Of course? Does it say anywhere that AI tools don't qualify? No?

If you can look at Midjourney's nearly 15 million users and say that AI tools are not promoting the progress of useful arts and tell me it's not promoting useful arts, you would be lying. There are 15 million users there that have had a mix of art experience and use it in their workflows, use it to create art that no one else would have otherwise, and get introduced to other forms of art. The number of users of this single tool rivals the number of artists employed worldwide across all art forms. *This is just one tool that I am pointing to*, there are numerous others, with millions of users or more for each of them.

These are tools that massively lower the bar to get into art. They massively lower the bar to create incredibly interesting art pieces. These tools empower artists to bring their dreams and ambitions to life when they couldn't otherwise.

This question is like the people in the 1990's questioning the utility of the internet.

22. Can AI-generated outputs implicate exclusive rights of pre-existing copyrighted works

They could if the subject is copyrighted probably. If I render an image of SpiderMan that's probably still infringement.

Also if I was literally trying to describe an exact image and reproduce it as closely as possible to the point it's not easily distinguishable.

But both of these come down to the use of the tool, not the tool itself.

23. Is the substantial similarity test adequate?

Sure. Trying to invent further ways to identify infringements would be tantamount to copyrighting style.

24. How can copyright owners prove the element of copying

They can't, again that's the point. Again that's a reason you can't copyright style, among other things. Again, you also can't prove a model *wasn't* trained with some data.

25. If AI-generated material is found infringing, is the creator of the model or the user responsible?

Typically the user. Unless it could be demonstrated that the model can reproduce the work that was copied with simple inputs a user would be likely to do during regular use of the model in such a way the user could unintentionally infringe.

25.1 Do open source models raise unique considerations?

Not really, they'd be pretty similar.

26. If a model is trained on material with copyright metadata, how is that copyright data propagated?

Not in any useful way, that's for sure. That'd be like telling someone to list all of the images they've seen that contain clouds. The individual usefulness of each copyright info would be next to zero and you'd need to list thousands of notices. This is assuming you could even trace that. Which you can't in any viable way. So you'd need to just provide all of the notices for all training data.

But you also wouldn't list all of the copyright information on the images that you used as reference as an illustrator, so maintaining the integrity of the copyright information is clearly not happening at a certain point. And that point is when it's transformed into something that isn't substantially similar.

27. List any issues you think policymakers should consider

Not thinking of any at the moment

28. Should content be labeled as AI-generated

Should my convincing photoshop job of Obama and Trump playing tennis together be labeled as photoshopped? Should my render of a flat earth in Blender be labeled as a 3D render?

No?

Should my render in Blender of a nice mountain view be labeled as a 3D render, what if it looks like a real photo?

No?

Why would AI-generated things be treated any differently from any form of digital media up until now? It would be against our best interests to do so since it'd just lend credibility to people who keep posting photoshopped things that are misinformation.

28.1 Who would be responsible for identifying AI-generated works?

N/A

28.2 Are there technical or practical barriers to labeling?

Yes, why are we labeling stuff? Is it to curb misinformation? Is it to put artists who don't use AI on a pedestal? We can't reliably detect AI-generated content without significant false-positive rates.

What constitutes AI-generated to a user? What if I just use AI tools to make a sketch which I personally draw over? Does that count? What if I guess wrong, does my account get banned?

28.3 What should the consequences be of mislabeling something?

N/A

This question is incredibly dystopian. "If thou shalt forget to label your AI-generated picture of a nice river scene, thou shalt be sent to jail for life"

Like what do you expect the punishment to be realistically? This would just be a chilling effect on all creatives if the punishment had any teeth.

29. What tools exist to identify AI-generated material?

One tool I know of is Hive <https://hivemoderation.com/ai-generated-content-detection>

There are many tools that claim to identify AI-generated material. Depending on the content they can have varying levels of effectiveness.

But I'll let you in on a secret. If all you are given is a few paragraphs of text, it's not possible to reliably identify AI content. And it will false positive on regular people.

<https://www.theguardian.com/technology/2023/jul/10/programs-to-detect-ai-discriminate-against-non-native-english-speakers-shows-study>

Oh also those big AI companies claiming to put watermarks in their stuff? That doesn't help since all of the open source systems are already capable enough and don't have watermarks.

Images too, can be easily circumvented, and can also false positive.

The end goal with all generative AI systems is to create content that is both high quality and indistinguishable from their subject matter. Identification is already proving impossible. It will only become less viable to the point of uselessness in the coming years.

30. What legal rights apply to AI re: likeness?

Pretty sure personality rights would still apply to things generated by AI... This question kind of indicates some need to clarify that since apparently that isn't self-evident to some people.

But also this is not copyright so I guess it's a given.

31. Should congress establish a new federal right re: likeness with AI

Uh, I think it would be enough to just clarify the existing law if it really is not that obvious...

32. Should style be copyrightable re: AI

????

I go and render some anime image that doesn't look at all like someone else's style. They claim it does. What happens? Anime all looks very similar. Does someone just own the ability to render anime images now?

What kind of standard do you apply here? Artists' styles can be very different even year to year.

If I train a model that's using a selection of AI art that *I have made using a model that has no input from a given artist* and then I select the ones that look the closest to the artist and say it's in the style of that artist, what happens? At no point has that artist contributed training data.

What happens if I train a model on some set of artists and say it's a mix of each of their art styles that forms my own style?

What happens if an artist claims that I'm generating artwork in their style, when I am both not generating artwork, and not using their style?

If I'm rendering things in a generic 3D style, can Pixar sue me? Just having that door open is an insane proposition which will have a major chilling effect at the very least.

These are questions that don't have sane answers, because the basic question is insane. Style is not copyrightable. That is non-negotiable. Style is far too vague and generic to make any sane judgements on.

33. How does the sound recordings section of copyright apply to AI-generated works

I'm not entirely clear on the question, but I don't think there would be much in the way of special considerations.

34. Any additional issues

AI tools will only grow in prevalence regardless of the decisions of the US copyright office. If you do not get with the times, other nations will force you to.

Some of the suggestions this document has made are deeply worrying for copyright in general. As someone who produces content that is used in training, and someone who uses the tools produced with that training as well, I am concerned if the copyright office is working in the best interests of either group. Many suggestions posited here would be greatly damaging to all creators. I hope that this letter helps find some understanding that prevents a chilling effect on all creatives.

Finally I hope to impress on you, any regulations you do to prevent training on artists' public content will not make them feel safer. Tools like Adobe Firefly will continue on improving and being able to emulate their styles no matter how difficult you make it at a policy level to develop said tools. The tools will not need to be trained on their content to emulate it accurately. Their styles will be emulated without invoking their name directly if needed as well. The value is there and this technology cannot be deleted. If you are hoping to mitigate some chilling effect in this respect, it is not possible to do so. You will only cause a greater, much further-reaching and damaging negative effect if you try to follow this line of thought.