Pierre Couillaud's comments to USCO on the AI and Copyright Public Opinion enquiry

For better readability, USCO's questions are in dark red italics, and my answers are in black. Questions that have been skipped are because I do not have enough technical or legal knowledge to answer them with confidence.

***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?*

Answer :

First, I must precise that I am not a professional artist, just an amateur of visual arts who follow a lot (over 1500) independent and freelance artists on social media (mostly drawing, painting, and animation related) and frequently communicates and share with them while enjoying their works. So I will talk on how it affected us based on my own experience and what i kept hearing for the past year from all of those artists and others :

So far, after one year since AI was forced upon us without our agreement, and without any of us having any desire for it, no one of us has seen any benefit from it.

AI is not a tool for artists, contrary to what the marketing of big AI firms says. It is a replacement. None of us have any use or desire for it.
Whether free lance or pro, all the visual artists and creators I follow are doing their job in a creative career... because they love to create (obviously). Why would they want a tool that automate themselves out of all creative process and removes all personal imput they have on their production? Of course they have no use or desire for it.

I myself, I follow all those creators and look at their work with bright wide eyes everyday because I love to see what they come up with personally, I love to have an open window on their imagination and inner worlds looking at their drawings.
Out of 1500 artists I follow and listened, less than 10 have started using AI or are enthusiastic about it; So it's about about 0.7% at most in terms of proportions, as far as I have seen.

In a broader sense, when I look, all over the world looking at the "benefits and risks", as USCO puts it, the only people in the entire past year I have ever seen AI empower to do things they couldn't do before are malicious actors.
The fraudsters that us AI media to falsely pass as artists, the phone scammers using other people's voice for impersonation, deepfake used for defamation, fake images used by extremist political groups as propaganda, sextortion on minors and children under 10 with fake pornographic images, Companies using unauthorised AI models to fire their entire creative staff, replacing them with models based on their work that plagiarise them without their consent. Etc, Etc...

The situation with AI as a whole, is that it does not really allow to do anything that couldn't already be done before: so unlike what companies that make AI systems try to claim in their sales pitch, it's actually **not** the same thing as a new medium, like photography or cinema when those appeared, for example.

The only thing AI does is replace.
Unlike Photography and cinema which opened new fields, all of the tasks we see AI algorithm do when they are introduced are not new things that could not be done at all before, but rather, already existing things that were done by people before.
There is not a single benefit in this to society :
The Creators lose their jobs, all of their income and worth re-directed to mega-billionaire companies, and the public pays the same price as before to now get inferior, mass-produced products of questionable quality and reliability. Certainly worse than when those things were done by human artists or workers before.
Everyone loses.

The only 2 groups who benefit from AI are :
-The giga-corporations who pursue hyper-profit at any cost. Now they can can cut jobs even more, flood the market with cheap and mediocre mass production, and also use AI as an excuse to reduce salaries of their remaining employees.
-The malicious actors of all kinds, as I mentioned earlier.

As far as the art job market goes specifically, even if in the long term, ML AI may not be an insurmountable threat to established artists with a reputation and follower base, but it will make it especially difficult for new artists to ever grow. All the low-skill entry level jobs for new level artists, where something only "good enough" is required, are the ones where AI will be used instead. And young talented people will not have any opportunity to grow big enough to be able to commercially compete with AI with reputation and skill later.

If this is allowed to continue unopposed and unregulated, the number of art jobs will be reduced considerably, and with it, the entire landscape in the creative industry will become considerably less diverse and vibrant.
Again, everyone in society lose from this, there is no benefit. Artists lose their jobs, spectators and audiences lose in that they are only given lower quality, partially or fully AI-made cookie-cutter images, movies, games, songs...

***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.*

Answer :

This excellent paper published in the Proceedings of the Association for Computing Machinery is a very good overview.
AI Art and its Impact on Artists :
https://dl.acm.org/doi/10.1145/3600211.3604681

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

Answer :

I cannot answer this question in much details as I am not a legal expert. But I will mention the obvious, EU AI act and specifically the dataset disclosure and curation requirements tied to it that are imposed on AI companies. I'll also say that yes, of course, international consistency is also necessary between different regulations, and worldwide laws will all align to each other eventually as far as AI regulations go. Because anything either internet-related or copyright-related go beyond any national borders, and always will always sooner or later become something that can be addressed only at an international level anyway, whether the different countries like it or not.

***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.*

Answer :

I do not have enough copyright law expertise to really answer this in details.
I do believe that actually, existing copyright and IP law is sufficient at least for the most part to tackle AI problems.

Virtually all of the currently existing AI models have clearly violated copyright law at multiple stages of their creation (at minimum 2 times : during training, with unauthorised usage of a copyrighted work for a commercial for-profit application, and during output, with unauthorised partial or total reproduction and distribution of a copyrighted work).

The current issue is more linked to the fact that the authorities, even after one full year, have stayed mostly inactive to stop this very visible infringement happening in broad daylight.
If only the current copyright laws were finally actually applied to AI systems, probably most of the problems would already go away.

The only argument I've ever seen AI companies make in defence of copyright infringement lawsuits is fair-use based. And fair use does not apply at all in the case of AI in my opinion (I will detail why in question 8).
But even if somehow fair use was judged to apply, USCO should keep in mind that fair-use only exists in USA law in the first place. What about those other hundreds of millions of copyrighted works from virtually any country on the planet, that are also massively included in the datasets of all current AI systems, without any consent or compensation then?

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

Answer :

I am not involved in this personally, so I can't answer in details.
I'll still say that, of course, it is largely an open secret that data sets for AI training are obtained by indiscriminate, mass internet scraping by way of crawler bots. There is simply no way to collect such massive amounts of data in a reasonable amount of time trough any other process.

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

Answer :

As mentioned in 6.1, the scraping is utterly indiscriminate with regards to source location, or nature of what is scrapped. So copyrighted images are used just as often as CC0 (without any consent from the copyright owner, of course, and without even notifying them).

There are very hardly any licencing models currently in use, but, much more importantly...
The important part is that the few that are in use, are nothing but a PR facade used by companies to achieve the outer appearance of rewarding contributors, without actually rewarding them in a meaningful manner.

I am sure many professional artists submitting their comments to USCO can give direct examples better than me on this, but still, even me: I know this indirectly. Because of the 1000+ independent visual artists I follow on social media, and I continually hear hundreds of them complaining about the impossibility to opt out in the first place of all those licences, and that hardly any of them have received any payment at all.

Not that any of those "licences" have any legal legitimacy of course: since they are one-way "agreements" that have not been signed or consented by artists.
For example, in the case of Adobe Firefly: the artists simply one day received an email from the company informing them that they have been signed into this "AI usage licence" from now on. They have not agreed to any of it. And are not being given an option to opt-out their images out Adobe's Firefly AI. (They can opt out of future models, but not out of the current one, that adobe has already commercialised and profits from).
Again, all of the above is what I heard from many artists I follow. I cannot personally testify it since I'm not involved in Adobe at all, but I see no reason to doubt them.

I suggest USCO can contact Karla Ortiz (twitter : https://twitter.com/kortizart) if they want a direct report from an artist of the Adobe Firefly example on licencing, since I have seen her talk about it often and have a conversation with an Adobe company official on twitter at one point).

There is also this article:
https://techcrunch.com/2023/09/30/how-much-can-artists-make-from-generative-ai-vendors-wont-say/

According to it, Robert Kneschke and 58 other stock photographers who have been forcefully subscribed by Shutterstock to the company's AI dataset program have compared the payments they received, and claim that on average, that payment is of 0.0078 $ per image.
Obviously I take this article with a grain of salt because I am not sure how trustworthy this journal is and I do not see it quoted many sources, but still, the claims align with what I heard many other artists say on social media, so I bring it to USCO's attention just in case.

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

Answer :

As mentioned in 6.1 and 6.2, Public Domain material is used indiscriminately, just as often as copyrighted material.
I do not think it happens very often to have specific training material created for the purpose of AI dataset because there is currently no punishment or limits for AI companies to scrape and use absolutely anything they want from the internet. So it stands to reason that they certainly don't want to spend any money commissioning specific material since they can get it all for free anyway. At least, for images and text.
Specific media like Sound or 3D models might be different here, but I do not know enough about this to answer.

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.*

Answer :

I cannot go far into the technical details as I am not an expert in this field. I have followed enough conversations on social media with scientists in the field of ML (Pr. Ben Zhao of University of Chicago especially, and others) to know the general idea with some confidence for image ML generators:

ML AI models for image generation work essentially as highly sophisticated collage machines, with weighted probabilities used to steer their output towards desired keywords (the prompts):

For images for example, they output by looking at thousands of images in their dataset of the "thing" they were asked to draw (the prompt), and then, the algorithm will try to guess what the image should look like, starting by taking a small element from one of those dataset images, and copying it on the frame.
Then, it will try to predict what should go next to this element that was just drawn, again by comparing it with the many images of the "thing" in the dataset, and copying in the next place another element from another one of the images that is statistically often associated with the bit that was just copied before. Then, it just keeps going like this until the frame is filled bit by bit.

All generative ML algorithm works under this principle of "guessing the output bit by bit",

regardless of what type of media it is.
It's the same principle again for text, for example: they start from a randomised word based on what the user typed, and then they try to predict the next word by looking in their text dataset what are the words that frequently follow the previous word. Then they keep going like this: by simply trying to guess one word after another based on what words often went together in the dataset.

All of the above, of course, means those algorithms are fundamentally, functionally not capable of creating anything truly new, they work only by re-arranging together various elements of their massive datasets into a collage-type patchwork output. Sometimes very, very complex patchworks, made of millions of individual elements taken from millions of different individual images from the dataset.

This paragraph above in particular seems particularly important to me to always keep in mind with regards to questions of copyright around ML image AI: they cannot create, only repeat, only output the contents of their database in a patchwork collage.

Of course, to output a convincing image purely made out re-arranged pre-existing parts, the number of potential usable parts need to be colossal, so that for any given detail of the output image, there happens to be an element somewhere in the dataset that coincidentally would fit well in this place. This is precisely why AI image generations systems need hundred of thousands of images before they can even output something recognisable at all, and billions to output something potentially good-looking.

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?

Answer :

I do not have enough precise technical knowledge to answer this.
Though I have followed on social media brief discussions between Pr. Ben Zhao of University of Chicago and other ML AI scientists on this topic, with them essentially all agreeing that this question is currently unresolved, and very complicated. (As in, still not certain whether it is even possible or not. And if it is indeed possible, it will certainly not be possible in a reliable manner anytime soon, not until many years).

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?

Answer :

No.
Without direct access to the dataset for checking, it is impossible to absolutely prove usage of a given piece of work. This is, of course, precisely the reason why AI companies absolutely DO NOT want to disclose their datasets, and are fighting with all their might against it. Without a direct, positive evidence, it is much more difficult (and risky) for someone to start a lawsuit, after all.

Without dataset access, the only indirect method that can be used to infer (not prove) usage of a given work in the dataset, is to try to coax the AI model into spitting out output that resembles the given work, either by using the name of the artist (which is often banned/flitered as prompts in

recent image models, for example, so not very easy) or a combination of keywords losely ressembling the work.
This is a very unreliable method that is time consuming, requires to have purchased the AI model in the first place, and cannot absolutely prove usage.

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

Answer :

Never.
Let's discuss the 4 main prerequisite for fair use and see why when it comes to generative AI, it can be shown that 3 out of 4 will not be passed in most cases, or in almost every case. (Whereas the last one, the criteria No3 are, in my opinion, a special case).

1. Purpose and character of the use, including whether the use is of a commercial nature or is for nonprofit educational purposes :

This one is very straightforward: any for-profit commercial AI system currently on the market, being sold for money or as a subscription model, do not pass this criteria. Whereas ML AI research and development projects, according to the law, might succesfully pass this condition.

However, it is still worth considering that even for ostensibly non-profit research projects, there have been cases (such as OpenAI to only cite a really big one...) where a supposedly non-profit project was suddenly turned around and commercialised overnight.
I do not have myself enough legal knowledge to asses the legal implications of situations like this, but this certainly raises questions about whether or not this condition for fair use is really passed by non-profit project if they are set up to be able flip to commercial and immediately start turning profits overnight...
I also make the argument in question 9.1 that even if a project is not directly for-profit, the makers of the project still benefit indirectly in many ways (including financially) from their project being successful.

2. Nature of the copyrighted work :

The vast majority of the ML AI systems out there that are likely to pose problems with regard to copyright in the first place will NOT pass this criteria.

Because they are designed to reproduce and output works that are very highly expressive in nature : especially image AI, animation AI, and sound AI (for voice and music).
This is not a coincidence of course, because those types of media are often the most valuable and sought-after commercially, so a lot of the works or sources in this category are copyrighted (and thus, many of those original copyrighted works end up being used in the datasets of those particular ML AI systems).

Text AI may be an exception to this only from time to time, for example, chat-GPT being used by someone to output a mundane list of groceries, or to help format an Excel spreadsheet, etc... this definitely does not reproduce artistically relevant and expressive work in this case.
But then again, chat-GPT being used to output a dialogue between 2 characters, or a novel, or the lyrics to a song, or a recommendation for colors that go well together according to dress code, etc

etc... all of those would NOT pass this criteria either because they are definitely expressive and/or artistic.

So even in the case of text AI, it is case-dependent and will not always pass.

3. Amount and substantiality of the portion used in relation to the copyrighted work as a whole :

I argue that, for this particular criteria, the traditional, pre-existing *fair use* law is NOT adequate to properly represent and assess the usage situation when ML AI is involved.
My reasoning for this is as follows:

What the traditional *fair use* law says for this criteria is that the quantitative and/or qualitative proportion of the original work that was used in the derivative should be assessed, and the more of the original work is expressed in the derrivative, the less likely fair use is to be concluded.

This works fine in the traditional case of a real, human creator taking inspiration or reference from an existing work to draw a modified copy of it, or a derivative, because the human artist drew his image according to the precise idea of it he had formed in his head.
So if there is similiarity between the 2 images, this, in turn, means that there was a substantial association and similarity within his head of what he wanted to draw and the original image (thus proving either reference, or plagiarism).
Humans draw with an image in their head : the idea of an image, that they can see in every detail mentally , and that they then put on a medium. They have a clear idea of what they "want" to draw.

However, the subtle problem here is that ML AI do not output their images in remotely the same way as how human draw. Unlike real creators (humans), who have a single and clear mental image of the work, and then put it to form on the medium, ML AI, on the other hand are only probabilistic collage machines.
ML AI algorithms do not have or follow any precise "image", they draw by looking at thousands of images in their dataset of the "thing" they were asked to draw (the prompt), and then, the algorithm will try to recreate an averaged, generalized view of the "thing", with some randomization introduced to it based on probabilities and randomized weights to avoid the output always being the same.
The resulting output is a generalized but also randomized collage-work of thousands of tiny parts taken from of all the images representing this "thing" that were in the dataset.
Even when the original work WAS, in fact, used and copied in part or in whole by the ML algorithm, and contributed, possibly very significantly, to shape its output image ; this output image could STILL hardly ever show any substantial visual similarity with the original work : because the process also involved avereaging and blending together thousands of other bits from other images in its dataset.

Thus:

Because of all of what I explained above, when the image was output by an ML algorithm, the lack of substantial similarity between the original work and the output image is NOT a proof that the original work has not been used, in whole or in part, by the ML algorithm to produce its output.
It also does NOT prove that the original work has not significantly contributed to, shaped, and influenced the output image.

My opinion is that when ML AI is involved :

Criteria No3 for fair use should be either ignored completely, or replaced by something more relevant that looks at the AI output as a whole and takes into account the thousands of database images that contributed to it.
It is pointless to try to make a 1-1 comparison between the output and a single specific picture from the dataset, because this is something relevant only to how humans draw images. ML AI does not output images like this. Trying to apply human logic to objects that do not follow human logic will not correctly represent the relationship between the original and the derivative.

4. Effect of the use upon the potential market for or value of the copyrighted work:

This one is rather clear-cut and straightforward.

I don't think it's very hard to conclude that the ability for anyone to produce copies of a copyright holder's work instantly and at no cost is clearly financially impacting the holder in a negative way. This becomes a lot more relevant when one consider the ability is not just to directly copy, but to create a huge number of slightly altered copies, as this further removes the need for someone to hire this artist to tailor-make exactly what the client needed.
I talked about this aspect a lot more a little later in question 8.5

In the case of research-only projects for non-profit and not made available to the public, on the other hand, this condition is probably passed for fair use.

However, if the non-profit project is made available to the public in any way (even for free), then once again, factor 4 is not cleared. Because then the public would have access to the AI model plagiarizing the artist, all the same as in the for-profit situation.

8.1. *In light of the Supreme Court's recent decisions in Google v. Oracle America [41] and Andy Warhol Foundation v. Goldsmith, [42] 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, [43] raise different considerations under the first fair use factor?*

Answer :

i do not have anywhere near enough legal expertise to answer this question.

I just want to leave a small comment about the pre-training vs fine-tuning topic :
I will argue that they should not raise any different considerations in terms of copyright or fair-use. Because ultimately, both fine tuning and pre-training are reflected in the collage output of an ML AI system.
Besides, even non-fine tuned models are also perfectly commercially viable and capable to repond to a lot of prompts and output good-looking images, so I do not see why factor 1 of fair use should be considered differently between pre-training and fine-tuning.

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

Answer :

I believe it is only common sense that companies that make datasets available for others to use for ML AI should be responsible to ensure the dataset is clean of any copyright or legal infringement and is of quality and relieable.
If that is not the case, they should be open to legal pursuits and be held accountable.

If they cannot be held accountable for making available (even for free) a compromised dataset, isn't it essentially just allowing the re-sale of stolen goods ?
Isn't it also essentially just this company selling a product unfit for purpose and not being open to accountability for it ?

The dataset-seller company should be open to lawsuits from both from the owners of the copyrights or other laws that were infringed, and also from the AI company that might have unkowingly purchased a compromised dataset from them.

***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?*

Answer :

I will just say that even AI research projects that are not for commercial use still do not clear the fair use bar in my opinion. At least, certainly not in every case ; as I have talked about in question 8 (especially the fair use condition 4. paragraph).

***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?*

Answer :

Because I have seen on social media various posts of AI users and makers, and discussions between ML AI experts (once again, Pr. Ben Zhao of University of Chicago) on this topic, I can say confidently that when it comes to images :

The bare minimum to have a functional model that can output recognizable (if quite bad) images is many tens of thousands, hundreds of thousands is more typical. Many millions is required for a model that can produce something at least decent-looking. Billions for a model that can output good-looking things and answer a large variety of prompts.
(I will not comment on media types other than images, because I do not have enough knowledge on those other cases).

When it comes to the fair use discussion, I will say that it does not matter at all.

I want USCO to carefully ask themselves the following question :

If using billions of copyrighted works modifies the fair use argument and makes it acceptable,
Isn't it the same as saying "if I stole 1 dollar from a billion people, I'm now a billionaire, but it's not really stolen money because it was only 1 dollar per person, it wasn't so bad."

In either a legal or common-sense perspective, since when exactly does scale justifies theft?
Theft is always theft, big or small.

***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? [46] 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?*

Answer :

Regarding how the effect on the market should be measured and verified, I do not have enough expertise in policymaking or those types of economical questions to be able to answer.

Regarding the second question however (impact on individual work or general work class) i think it is very relevant for USCO to be bringing up this topic : because there is a fundamental difference between market impacts of AI systems vs other non-AI related usages.

And that difference is **iteration :**

using images as example:

In a classic usage-case of a copyrighted material, where no AI is involved, we are generally only talking about re-using or reproducing the copyrighted image in an exact manner. (Mostly so : as even in cases where the original work is used by someone else as the base to create a derivative work, usually the number of different derivatives involved will still be reasonably small).
Therefore, the commercial impact of the usage, or the derivatives, will probably be limited only to the specific work used, or, at the absolute most, to the specific artist who made this original image, but almost certainly not affect other artists in the same field or category.

But when we are talking about usage of a work for ML AI dataset, suddenly, we are talking about the ability to not just reproduce identically, but also, if desired, to iterate almost endlessly, and very quickly, on the original work : to create a near endless amount of different modified derivatives of the work, instantly.
The important part is that this means the derivatives can be fine-tuned to fit more on what the AI user wants, thus reducing much more the need to go back to the original artist and considering hiring them to make a more closely tailored version of their art, for example.
And then, if only about a dozen or so of a given artist's work are involved in the dataset, the model will probably be able to plagiarize quite consistently this artist's entire artstyle. Now drastically reducing the need for anyone to hire this artist at all, because they can have the ML AI draw almost anything they prompt with a "close-enough" version of this artist's artstyle.

And finally, in some cases, this certainly can go as far as impacting other unrelated artists working in the same field too. Especially if an artist that is particularly skilled or popular ends up having their artstyle style plagiarized by an AI model, many users will not even look anymore at other artists that draw a similar category of works or topics, because they can use the AI model that plagiarized the best of the best at it instead.

So i will say the market inquiry when usage of ML AI is involved should definitely be considered at all 3 levels : specific work, body of work of an author, entire class of topic/artstyle.

I would argue that the usage of a single work of a given author for ML AI will inevitably have consequences and compete at least up to that author's entire body of work. (Again, due to the ability of ML AI to iterate so many different derivatives instantly).
Impact on the entire general class of works is, I would suspect, variable on a case by case basis.

I will finish by saying that all of the above should always be remembered and taken into account very carefully when assessing the commercial impact on the original artist.
Although a subtle difference, the ML AI's ability to instantly reproduce countless different derivatives is massively influential. It means the impact of usage of a copyrighted work for ML AI dataset is inherently always much higher than non AI-related cases.

***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)?*

Answer :

Opt-in only.

Seeking and obtaining consent from the author/copyright owner in a legally-binding contract should be a requirement for usage of their work in a dataset. Because it is a part of copyright law that the owners of the copyrights and makers of the work should have the right to decide how their work is used and for what purpose. Plain and simple. Fair Use does not apply to usage for AI dataset, as discussed in question 8. and 8.1
Using copyrighted materials for AI datasets without consent constitutes Copyright infringement and IP infringement.

Opt-out will never be a viable solution for 3 reasons :

Firstly, it is functionally not feasible for artists : there are thousands of AI stratup companies out there, and beyond that, tens of thousands of isolated individuals doing their own "open-source" models.
Now, I want USCO to try to imagine themselves in the shoes of an artist : and imagine how much time it would take to scour all over the 4 corners of the internet for all the datasets of every single one of those AI models, one by one to try to find out whether one of their works has been used in the dataset or not... (most of those individuals or AI companies will not even accept to disclose their dataset altogether, leaving the artist with no choice but to try to use their own name as prompts to see if the model might have their work in it or not, but not providing definitive, legally-binding evidence).

Then, on top of this, I want USCO to imagine how much time it would then take to fill all those tens of thousands of opt-out forms and gather and send all the necessary documents for them, one by one...
Even if artists dedicated 5 hours a day to it every day, they probably could not even keep up with all the new models being released every week, and so the list of models they need to check and opt-out from, would only keep getting longer over time rather than shorter...

This is why **Opt-out is a farce**. It is simply not even materially possible.

Secondly, opt-out is only possible for their future works and on future versions of a given model. The currently-marketed models already using the works of the artist without consent cannot be made to "un-learn" the work that are already part of the dataset after it was trained.

Thirdly, in a more general sense, because it is not fair that the burden of work should be put on the artists to defend themselves.
The ones profiting from the AI models are the AI companies and individuals making them, so the burden should be on them to spend the necessary time to seek and obtain consent from the copyright holders.

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

Answer :

All usages.

The problem with legalizing IP infringement for the sake of research is that it becomes very easy (and financially tempting) for AI companies to pull a bait and switch where they begin as a "research only" project to have access to any data they want, improve and refine their AI system enough to make it commercially competitive, then, simply switch out of research , into a commercial product.
This bait and switch is, in fact, precisely what Sam Altman did with OpenAI for example, with the Open AI company originally starting, and operating for years, as ostensibly non-profit. So it's not a theoretical problem, it's already factually proven that this is what't happening, and what companies will absolutely continue to do.

The other, more general reason why consent and licenses with copyright owners must always be obtained, is because it has always been customary, in scientific research culture, to reward participants in research projects in some manner. in recognition for the value they add to the project. And AI projects should not be exempt of this rule.

I can speak of this personally , because i have an education in science (i have a master's degree in Paleontology and evolution). And, although this is in fields different from mine, I have friends and acquaintances from my university in psychology and medicine, and sometimes, when they did simple projects such as requiring volunteer participants to play a game where their choices and reactions were analyzed for the study, they would always reward each volunteer with something like 50-500 euros depending on the time involved for their participation. I would also see on the campus sometimes recruitment posters for volunteers on small studies like this, from various teams in the university. It's not much money for the university department, but it helps to get more participants, and more importantly, it acknowledges that the volunteers brought something of value to the project, and made possible a study that would not have been possible without their participation.
I think this is fair.
I think this is how it should be.
AI systems are worthless and cannot do anything without a database.
I do not think it's fair that people who contributed to this database should be ignored.

Even if the project is not for profit, the scientific team still hugely benefit from it in terms of prestige, university funding for their department in reward to their good work, scientific article quotations from peers, etc... the dataset clearly brought something of value to them.
Why should they be entitled to profit from the works included in the dataset without giving at least something back to the dataset contributors ?

And finally, in a more broad sense, you don't just enlist someone in a scientific study without their consent.
You can't test a medical drug on someone without their consent.
You can't conduct a psychological analysis of someone and publish it in a paper without their consent.
Doing this in secret, behind their back, is both deeply unfair and deeply disrespectful to them.

Why should AI corporations be allowed to enlist the entire body of work of an artist in their scientific project without their consent? Or use someone's face or voice and replicate it without their consent ?
It's not just data... to those people, it's their life's work, and their identity. It's dehumanizing, unfair, and disrespectful. This is not how people should treat each other in a fair society.
That's all I have to say about this.

***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?*

Answer :

No.
Because using a copyrighted work without consent should be illegal in the first place. (question 9.)

***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?*

Answer :

As I have argued in question 9: the AI developper should always seek and obtain consent trough a legally-binding license with the copyright holder.
However, i agree it is certainly not realistically possible for companies to sign and handle tens of thousands of individual contracts with all the copyright holders, one after another.

The solution to this, in my opinion, is to pool the many AI usage licenses into trusted datasets :

The copyright holders who want to license their work for AI training will sign a license with a dataset-vendor entity, whose entire purpose is to handle all these contracts in batches and pool these licensed works to create trustworthy and ethical datasets.
These various datasets can then, in turn, be licensed to AI companies.

This dataset vendor entity can be a standalone commercial company specializing in this, or a non-profit company set up by a guild/group of artists to manage AI licensing, or a government-staffed agency, or a separate branch of an existing AI company, it really doesn't matter.

The only thing that matters, is that there is careful oversight from independent observer to ensure the datasets are indeed certified trustworthy and only contain licensed or CC0/public domain materials. And to enforce that the company is not using various means of pressure to strong-arm creators into signing contracts they do not want to sign.

One small thing that I will add is that there should be a clear specification in the license contract as to a broad categories of projects that this work is being allowed to be used for. So that copyright holders can still retain controls as to what kind of AI projects and applications they want to allow their work to be used for.

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?*

Answer :

If a copyrighted work is used without the consent of the copyright holder, then this simply constitutes copyright inringement/ip infringement ; and the normal laws for this type of crime should apply. Plain and simple.

In addition to fines and punishment for the infringing entity, the entire dataset that contains this non-consented work also becomes immediately invalidated, and cannot be used anymore for AI training. (even if this is just a single work.)
Furthermore, all AI systems that were trained in the past (even just in part) with this polluted dataset, and all media output by these polluted AI systems, are also to be deemed legally invalid, and not used commercially anymore. Pulled from markets, if they were already being commercialized.

The above paragraph only consists of nothing more than the normal application law around copyright infringement :
Let's take the example of a movie that was released in theaters, and it is later proven that, let's say, the costume design and appearance of one of the characters was taken or plagiarized from another work, whose copyright owner has not consented for its use in the movie.
In this case, unless an agreement is reached between the copyright owner and the studio, the **-entire-** movie would be invalidated and pulled from theaters unless the costume design and appearance of the character is altered by the studio to remove the copyright infringement.

This is how the existing law works, and it is fair, it protects copyright and IP of creators.
And this is how it should work with AI datasets and their output too.

Note that, in the case of a polluted dataset, another acceptable solution could be for the AI companies to have all the affected AI system "unlearn" the specific entry in the dataset that is copyright-infringing but keep the rest. However, this completely hinges on whether it is even possible or not for ML AI system to unlearn anything. Which, as stated in question 7.3, is largely a scientifically unresolved question at present.
And even then, of course, if successful unlearning is made, all of the older output media made by the affected AI system before unlearening (when the problematic copyrighted material was still a part of their dataset) are all still legally compromised, and must still be pulled from market.

Finally, it should be noted that simply censoring specific keywords is not good enough in the case of a polluted dataset, either. (current AI systems sometimes do this : for example, : preventing the names of specific artists to be used as prompts for images).
The reason this solution is not satisfactory is because there are plenty of other ways people have already worked out to still coax the AI model into spitting output related to the artist's work without using their names, for example, by entering text prompts that precisely describe one of the artist's works.
Also, even if the specific artist's name is not used as prompts, the work will still be used in other ways potentially in any other images output by the model (for example, if artist A has a painting of a cow in the dataset, if the user asks the prompts of "a meadow with a cow", there is no way to be sure whether the AI model used the copyright-problematic cow painting of artist A as one of the materials for the output image, or another unrelated, non-problematic cow image in its dataset).

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?*

Answer :

First of all, i will say that this very problematic situation can probably be avoided most of the time by requiring that any contract that involves copyright transfer must have a very clear clause on how the work will or will not be used for AI applications and detail those applications precisely.

Still, if a disagreement still arises between the creator and the copyright holder, i think this is something that should be judged on a case by case basis in a lawsuit. I can see too many different situations for a blanket-rule on this to always be fair for everyone.

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

Answer :

I already answered this in question 9.3.
Again, the best method in my opinion is to pool all the usage licenses to an entity that specializes in managing those contracts in batch, and then create datasets for AI training that can be certified to be clean and ethical. (See question 9.3 for more details).

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

Answer :

I think that realistically, it should be feasible for all sectors, if the licensing company is specialized in setting up those kinds of contracts in large batch and managing them in groups, and has the resources and experience dedicated in doing so. (Again, see question 9.3 for more details).

The company is simply there, ready to accept applications from any artist that want to license their work for AI training. If the artist is not OK with the offer they are free to walk away, and if they are OK with it, then the contract signed is the same for everyone (thus facilitating management of the huge number of contracts).

10.2. *Is a voluntary collective licensing scheme a feasible or desirable approach?* [49] *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?*

Answer :

Yes, again, as explained in the many questions above, i think a collective licensing is a good approach, and in fact, is the only realistically possible approach given the massive number of contracts required. (Again, see question 9.3 for more details).
And that the entity that manages all those licenses should also be the entity that then creates the datasets from the licensed works, so that those datasets can be certified clean and ethical, thanks to independent vetting and rigorous oversight.

The status of antitrust laws and other things mentioned in question 10.2 are far beyond my personal competence in economy so I won't answer them.

10.3. *Should Congress consider establishing a compulsory licensing regime?* [50] *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?*

Answer :

No.
Again, any kind of usage of a copyrighted work should require consent. (as I already stated in questions 9. and 9.1).
Because it is a part of copyright law that the owners of the copyrights and makers of the work should have the right to decide how their work is used and for what purpose. Plain and simple.

10.4. *Is an extended collective licensing scheme* [51] *a feasible or desirable approach?*

Answer :

No.
Again, any kind of usage of a copyrighted work should require consent. (as I already stated in questions 9. and 9.1).
Because it is a part of copyright law that the owners of the copyrights and makers of the work should have the right to decide how their work is used and for what purpose. Plain and simple.

Besides, such a setup seems to me to be extremely ripe to cause discord and disputes between the creators and the collective license holder who took those important decisions on their behalf.
Plenty of potential conflicts of interests for whoever directs the collective license, plenty of influence struggles to control how work is used ; strong commercial incentives for AI companies to bribe the collective license holder to give them favorable deals, or undercut concurrent AI companies, etc,etc...

Overall, a Collective Licensing seems to me like simply begging for all kinds of trouble to happen, and I don't see a single benefit in it, either.

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

Answer :

I don't feel I have enough knowledge of the different areas of copyright laws to answer this question.
Realistically, I can see somewhat different licenses being required for a soundfile or an image, or a book, due to the different mediums, but I'll let artists who work in those fields answer this better than me.
Regardless of the type of work or media involved, however, the license should always be based first and foremost on consent of the copyright holder, of course.

***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)?*

Answer :

I already answered this in question 9.3 and 10.2
Essentially, again, in my opinion, it does not matter who is managing those collective licenses, as long as they are subject to very close scrutiny by independent observers.
To ensure they do not strong-arm artists to sign contracts they do not agree to, and that they are properly building the datasets using only the licensed works as stated in the license contracts.

There could even be more than one entity proposing different license contract offers, for different types of medias or applications, etc... rather than a single monolithic entity.

Why not ? As long as all those licensing entities are all subject to the same careful monitoring and accountability, to ensure they play fair, and given in return the official "seal of approval" that the datasets they offer are clean.

***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.*

Answer :

I do not have the technical knowledge to answer this.
But I can give the general answer that much like the "unlearning" of a particular dataset entry, this seems to be currently an unresolved question amongst experts in the field. (According to conversations I followed on social media between Pr. Ben Zhao of University of Chicago and other ML scientists).

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

Answer :

I do not have enough knowledge in marketing to answer this in detail.

I would like to just leave 2 comments on this topic :
First : regardless of the economic impacts of companies, a licensing requirement is what is fair and necessary to put in place to upload copyright laws.

Second : enforcing a requirement to use the vetted ethical datasets and forcing those datasets to be fully open to public is very likely to have a beneficial for the general market of AI systems because it will reduce monopolies and give a more even ground to small and medium companies.

This is because if everyone has to use the same datasets and make them open, one of the key competitive advantages that gigacorporations have is forcefully evened.
Due to their much bigger acess to computing power, general internet information and money, giant companies that currently already have an in-effect monopoly over tech are capable to gather works and data from all over the internet far more quickly than a smaller company, and only keep getting more ahead in the field of ML AI too.

On the other hand, although I am certain those companies would all immediately proclaim this to be "anticompetitive" and an "unfair advantage to their competitors", it is not true.
Because it only takes away one of the advantages they have, but they still have an advantage in computing power over smaller companies (something also essential for AI development).

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

Answer :

There is one extremely thorny potential situation I can see certainly arise in the future that I have never seen anyone discuss anywhere before, and I really would like like to bring it to USCO's attention :

Essentially, as I stated many times by now, the correct setup in my opinion is an Opt-in only system. Where usage of copyrighted works for AI datasets is forbidden without the copyright holder's consent, obtained in a legally binding license.

However, there is a very serious flaw, even with this :

If an AI company wants to appropriate themselves the artstyle of artist A , who is refusing to license his work for any AI usage ; there is nothing that technically prevents this AI company from hiring another artist B, to create works in the style of Artist A for them to be used as ML AI datasets.

There are many people who are incredibly talented at doing impressions of other people's voices.
There are many people who are incredibly talented at drawing in the arstyle of others.
Etc...

The AI company would not have broken any laws, since the works are all consented and licensed by

artist B.
and Artstyles are not copyrightable, so artist A would hardly have any recourse against this...

And yet, this is clearly not fair.

The only straightforward solution to this I can think of would be to grant copyright to artstyles. However, I do not think at all this is something realistically feasible, nor desireable.

What exactly is a distinct artstyle and how to gauge at what point it is distinct enough to be copyrightable, or when is it "too similar" to another... all of those steps are far too subjective, all of it would need to be judged on a case by case basis, this would just be an inextricable tar pit.
Copyrighting artstyles would also have dreadful ripple consequences on fair use and secondary creations, which, I believe, would end up being detrimental to artists worldwide and stiffle creativity.

However, how to resolve the "AI training through proxy artist" problem explained above then?"

I could not come up with a solution to it. Hopefully USCO can.

***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?*

Answer :

Yes, of course.
As I already mentioned indirectly in question 10.2 and 11, I argue that Datasets should all be made by one or multiple specialized entities : they sign and manage the usage-agreement license from artists that want to license their work for AI training. And then collect these licensed works into thematic datasets for relevant purposes. These entities can also make on their own other datasets from CC0 materials if they wish, of course.

All of those entities and the datasets then create must be closely vetted and monitored by independent observers to ensure they play fair and do not strong-arm artists into unwanted license deals, an also make sure the datasets are clean and ethical.
Even the CC0 datasets must be subjected to the same oversight, to ensure no copyrighted works are added (accidentally or maliciously) to them.

When it comes to the companies that make AI, they should simply purchase licenses to use those vetted and greenlit datasets from the dataset entities.
No AI company or individual should ever be allowed to train an AI system on a non-vetted and non-greenlit dataset. This shall be illegal without exception.
The reason for this being, because of the massive amount of data involved and complicated collecting process, it is quite simply impossible to prove a dataset to be clean and ethical if it has not been continually observed and documented through its entire creation and growth (the purpose of independent oversight on the dataset entities).

If an AI company does not wish to use one of the greenlit dataset from one of the dataset entities, for whatever reason, they are, of course, free to create their own dataset by themselves.
But in this case, they must be subject to the exact same oversight and vetting process as the dataset entities : the whole process from the first licenses being signed all the way to the dataset being

completed must be independently observed and vetted, and the AI company must provide public access to the dataset for anyone to review, otherwise, they should not be allowed to use this unproven dataset.

15.1. *What level of specificity should be required?*

Answer :

The ability to browse the entire dataset and consult every individual entry in it.
All the information about the author, copyright status, and other relevant information must be kept linked with each entry. Where was it obtained from, when, and trough which means.

In the case of entries of copyrighted materials, it would also be good to keep linked to each of those entries the relevant usage license agreement between the copyright owner and the entity that is making the dataset. (I am not sure how feasible this is from a technical point of view, but it doesn't sound to me like something impossible to do).

15.2. *To whom should disclosures be made?*

Answer :

Everyone. At all time. Without restrictions. Without exceptions.

Totally public-open dataset, available on a certified registry website, for anyone to consult, and browse the entire dataset at their will and check any media of any kind present in it.

Access to those clean-certified datasets should be hosted and provided, once again, by the entities that are managing the collective licenses of copyrighted works for AI training.
(it is worth noting that to facilitate navigation, a search engine allowing to search entries in the dataset by keywords should be provided by the registery website. Similar to the already existing 3rd party website " haveibeentrained.com " which is used to browse the Laion5B database).

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

Answer :

This is a somewhat complicated question, but in my opinion, the best method to keep it simple is to simply require 2 things mainly :

First, that the AI developer should double-check themselves that the 3rd party model they are using is only using the vetted, certified clean datasets and fulfills all mandated safety requirements.

Second, to conduct risk assessments as to how the interaction of the 3rd party system and their own AI system might result in unexpected properties or potentially enable unexpected (and unwanted/dangerous) new usages from the userbase.

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?*

Answer :
I do not have the relevant industry knowledge to answer this.

I will comment that the cost of maintaining the records should come either from the AI companies or from the certified dataset-vendor entities (or both).
Ultimately, those 2 entities are the ones profiting from the AI systems, so they should be the ones to shoulder the cost of maintaining this necessary database, as an inevitable cost of conducting the business they are in.
It is both unfair and uneeded to offload this cost on public money or other 3rd parties.

I want to also remind USCO here of what I had already said in question 13 :
Enforcing a requirement to use the vetted ethical datasets and forcing those datasets to be fully open to public is very likely to have a beneficial for the general market of AI systems because it will reduce monopolies and give a more even ground to small and medium companies.

I think these are the most important points. There may be other things people who are more knowledgeable than me on this could add.

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

Answer :

Already answered in question 9. : usage of copyrighted material without consent from the copyright holder should be illegal in the first place.

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?*

Answer :

EU AI Act has strict requirements of complete dataset discolsure high-performance for most AI systems, and the US should adopt such requirements too as I already stated in question 9.3 and 10.2

Fully open dataset available for anyone in the public to freely review is absolutely indispensable for AI systems to even have a chance of being deemed trustworthy.

And as said in question 4. : international harmonization between regulations is both necessary and inevitable in the long term on any matter that concerns copyright or the internet anyway, as those are domains that always cross borders.

***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?*

Answer :

No, never.

Copyright law exist to protect human creativity and expression. There is no creativity and expression involved in an automated process.
And prompts do not represent a "creative contribution" because they are merely vague directions/instructions, they are not direct inputs with direct predictable and mesurable influence on the output.

The situation of the user of an AI system is never that of a "creator", but rather that of a "client" :
They are in a comparable situation to a client commissioning or hiring a human artist to do a piece for them.

When a client commissions an artist to make a piece for them, they also give the artist instructions, that can sometimes be extremely precise and require the artist to re-do or correct the piece dozens of times. Yet in spite of all those instructions, the copyright always goes to the human artists, with no contributions recognized from the clients, because ultimately, the client, did not have any -direct- input on the finished piece.

The only difference in the case of an AI user, is of course that unlike a hired human artist, an ML algorithm is incapable of creating anything new, and it is merely a collage-machine remixing together existing works, as explained in question 7.1.
Therefore, in the case of an AI system, no copyright at all should be given to anyone for the output : because it was purely generated through a randomized process, with no human being having any direct creative input over it at any point.

This question of whether or not outputs of AI systems should receive copyright was already discussed in good details by USCO themselves in the published letter below.
The arguments used by USCO are all professional, logical and excellent, i fully agree with all of the statements made in the letter. So to answer question 18, I will refer USCO to their own work:

https://www.copyright.gov/docs/zarya-of-the-dawn.pdf

The most important parts to the question are especially the paragraphs that go from page 8 to page 12:
"2. Application of Copyright Law to Midjourney Images"
"3. Images Edited by Ms. Kashtanova"
"Conclusion"

When it comes to mixed works (that are made of a mix of AI-generated parts and human-created parts), such a work cannot ever receive full-copyright, because it is not fully human-created. Only the parts of the works that were human-created can receive copyright.

This is exactly the conclusion that USCO also came to in the letter linked above:
In that case, total copyright was not given to the entire comic book.
Only the hand-made elements were granted copyright: the text in the speech bubbles (which was hand-written), and the layout of the speech bubbles inside the frames (because those were directly hand-made).
The images themselves were not copyrighted at all as they were AI-generated with only minor hand-made corrections that did not alter the nature of the images.

***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?*

Answer :

I do not have enough knowledge on the precise text, and lawmaking in general, to answer this.

I can just leave the small comment that, in my opinion, it would be a good idea to make it clear and visible to anyone seeking to register a copyright from the start what exactly the rules are :
ie : anything AI generated cannot be copyrighted.
Provide a visual example, such as :
In the case of an AI-image with a handmade correction where the arm of one of the characters was redrawn in a different position, the only thing that can be copyrighted is the floating, isolated arm on a blank canvas (only the hand-drawn parts), none of the AI ML generated parts.

***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?*

Answer :

No.

In the past year, we have all already seen something that can only be described as an utterly frenetical rush, on the part of almost every major company imaginable, to inject ML AI systems into almost every aspects of their products as quickly as possible...
I'm sure USCO can see just as well as me that there is already more than enough incentives as it is out there to encourage companies or developers to pursue AI development or to attract collossal investments.

I really, really fail to see any argument whatsoever why yet another incentive would be necessary.

Especially considering that providing this extra incentive (granting legal protection to AI outputs) would require some very far-fetched breaking of the existing copyright law to even exist.
See again my answer to question 18 especially as to why this is the case.

I also fail to see any argument why the standard, already-existing copyright protections for computer code would supposedly not be enough to protect AI systems.

***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?*

Answer :

already answered with more details in question 19. and 20.
Copyright is not deserved by auto-generated media that is only a collage of existing works from a dataset, and that do not involve direct expression of a creator's mental concepts, feelings and ideas.

The very idea of a " sui generis right" is nonsensical because, as already explained at great length in question 19 : the user of an AI system is not a "creator" but a "client".
As per USCO themselves in the letter quoted in question 19 : the vague instructions given by prompts are not a direct creative input and do not contribute in a predictable and mesurable way to the finished output.
Once again : in the case of an AI system, no copyright at all should be given to anyone for the output : because it was purely generated through a randomized process, with no human being having any direct creative input over it at any point.

***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"? (52) If so, how?*

Answer :

I do not have enough legal knowledge of the US Constitution to answer this question in detail. obviously

I can however say with total confidence that such protection for AI outputs will **NOT** "promote the progress of science and useful arts".

This is because as i have mentioned many many times by now, ML algorithms are inherently incapable to create anything.
They are only collage machines that remix together pre-existing material, they cannot ever "innovate" or create anything truly new.
How can something that do nothing but remix and collage the past be promoting progress ?
Real creativity and innovation will only ever come from human beings, from artists and scientists, as always.

As I already said in question 1 :
The more ML AI systems are used to displace and hijack the creative process away from artists, the more the diversity of the creative landscape becomes impoverished, the more everything becomes cookie-cutter, mass-produced and identical.

This isn't "promoting the arts"... this is the exact opposite : this is promoting the disappearance and impoverishment of creativity, imagination, innovation and creative culture.
Everytime someone chooses to use an AI system to output an image or a novel or what have you,

this is a lost opportunity for this person to express themselves, or to hire and contribute to the growth of an artist to work together with them in commissioning a piece.... Instead, that person chose to just had a machine sipit out a randomized, patchworked collage of other people's expressions.

This isn't a "useful science and art" either, as mentioned in question 1 and above :
Nobody benefits from this in society.
Artists lose their jobs, spectators and audiences lose in that they are only given lower quality, partially or fully AI-made cookie-cutter images, movies games, songs...

The whole past year has demonstrated perfectly that only giga-corporations and malicious actors benefit from widespread ML AI usage across all of society...

***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?*

Answer :

In fact, it is the exact opposite : it is impossible for the output of an ML algorithm to not implicate the copyright and reproduction rights of works used in the dataset.
This is is because, as explained in question 7. 1 ML algorithms are only probabilistic collage machines that inherently cannot create anything new.
Their method to generate their output is to stitch together thousands of minuscule elements taken from the images of their dataset, and arrange them in a shape vaguely matching a generalized and randomized view of the thing the user prompted it to draw. This generalized view is itself formed by avereaging together all the images of the dataset that represent the thing the user prompted it to draw.

Understanding this, it is easy to see that everything output by an ML algorithm, under every circumstance, will always involve the copyright and reproduction rights of the works used in its dataset ; since after all, the entire output is made only of a collage various small bits reproduced from the works within the dataset, some of which are copyrighted.

I am not familiar at all with the nature exclusive right specifically, so I will not comment on that.

***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?*

Answer :

While I do not have enough legal expertise to know all the ramifications of Substantial similarity in copyright law; I still want to make here the important following comment :

This question seems to more or less converge with what I discussed in question 8. specifically, the requirement No3 of fair use law : (Amount and substantiality of the portion used in relation to the copyrighted work as a whole).
As I said back then, my position is that trying to assess copyright infringement from direct comparison between the original and derivative work is not adequate when ML AI systems are involved, because the process of an ML Algorithm to output something is completely different to how human artists reference existing works:

When the image was output by an ML algorithm, the lack of substantial similarity between original work and output image is NOT a proof that the original work has not been used, in whole or in part, by the ML algorithm to produce its output.
It also does NOT prove that the original work has not significantly contributed to, shaped, and influenced the output image.

(See question 8 fair use criteria 3 for the detailed reasoning).

***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?*

Answer :

This was already answered in question 7.4 :
It is not possible to prove with certitude usage of a given work in an AI model without access to the database archive. At best, it can be inferred (not proven) by trying to get the AI model to reproduce the specific work through carefully selected prompts related to the artist or to this one specific work.

This is one of the many reasons why, as I stated in question 9.3 and many times afterwards , the datasets must be safekept by the same companies that are the managers of the licenses for copyrighted work usage, and the makers of those datasets (the dataset vendor entities). These entities must be continually put under oversight and careful scrutiny from independent observers to ensure they do not temper with the dataset records and maintain them properly (among other things).

This is also one of the many reasons why, as stated in question 11. and 15., those datasets must also be completely open for anyone in the public to consult and explore at all times.

***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?*

Answer :

Everyone responsible must be held accountable in case of infraction:

the user of the AI system who commanded the generation of problematic AI-generated media, the platform (website, magazine, application, TV channel...) that allowed the problematic AI-generated media to be displayed to the public, and the company (or multiple companies) that made the AI system and/or gathered the dataset.

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

Answer :

So-called "open source" AI models cannot be trusted under any circumstances, and must be outlawed altogether.
The lack of any oversight, verification and vetting process on their dataset makes them inherently impossible to ensure that they do not infringe on copyright, and in fact, nearly certain that they DO infringe on copyright, intentionally or not.

I can provide USCO with a very real example of just that, which is ongoing right now :

Please take a look at this training Dataset currently available for anyone to use on huggingface : https://huggingface.co/datasets/animelover/danbooru2022

(Hugging face, by the way, is generally considered to be one of the more trustworthy and reputable sources for training datasets in the open-source AI community. By all means, I encourage USCO to ask AI companies or workers what do they think of this site to verify it, if you do not trust me on my word.)

So what is the problem here? Well, the copyright status of the dataset linked above is filed by Huggingface as CC0-1.0 (as you can see under "licence"). Or in other words, supposedly Public Domain.
But, that's not the case. At all.
This dataset is made of virtually every image that was uploaded on the Danbooru website in 2022 with certain tags attached to them, as stated by the uploader in the comments at the bottom of the page. over 4 million images total.
(One of those specific tags the uploader selected to choose which pictures to add to his dataset is "3D", i will use this one in examples later, so keep it in mind).

So what is this source site, Danbooru ?
Danbooru is a well-known repost website for images and artworks that has existed for well over a decade. It is essentially a semi-pirate site that relies purely on unauthaurized reposts. Essentially, users of Danbooru copy/paste and then repost on the site any image that they want to repost for others to look at, collected from anywhere on the internet, or elsewhere.

As a result of this, **almost ALL images on Danbooru are, in fact, copyright-eligible, original works** : a lot of them, especially, are works from independent/free-lance visual artists who often post their works on twitter or their own portfolio websites (and from there, are uploaded to Danbooru by the website's userbase, without any notification or consent from the copyright holder/creator).
There is hardly ANY image in the entire Danbooru website that is CC0/Public Domain.

… But wait, it gets worse :
In addition to original works , there is also plenty of official Artwork from various companies or IP holders on Danbooru too : promotional materials, official artworks, artbook contents, in-game art assets ripped directly from videogame files with 3rd party tools, etc....
These can be recognized on Danbooru easily because they have the special tag "official art" attached to them on the website.

For example, take a look at this one:
https://danbooru.donmai.us/posts/5861716?q=official_art+3d+
This was uploaded to Danbooru in 2022(11-29) and it has the 3D tag in it, so it's in the Huggingface database linked above.
It's the official poster for the Super Mario Bros Movie...

Another example:
https://danbooru.donmai.us/posts/5402417?q=official_art+3d+
This was also uploaded to Danbooru in 2022(6-03) and it also has the 3D tag in it, so it's also in the Huggingface database linked above.
This is the Official artwork for the character Chu-li in the Capcom videogame "Street Fighter 6"....

I could go on with examples, but I think I made my point clear.
The -supposedly- reputable HuggingFace dataset repository website simply allowed a completely random person, whose identity can be known only by his account name of "animelover" to upload this dataset he collected himself, as CC0, without any verification or accountabilty.

Now, even someone with genuinely honest intentions to "ethically" train their AI system on CC0 material only, will have at least 4 millions copyrighted, non-consented images in their dataset and they are probably not even aware of the problem ; both infringing on the IP of all those copyright-holders in the dataset, and also unknowingly putting themselves at risk of lawsuits too.

I want USCO to be clear on one final thing here :

This example is NOT an isolated case.
This is very much the Standard Operating Procedure for the so-called "open-cource AI community".

Anyone uploads any dataset they want on some dataset-repository website somewhere, then everyone just uses them (or collects their own dataset), no one questions anything, no one verifies anything, no one takes any responsibility for what they do or what happens.

And this will NOT change unless a strict vetting process by an independent agency is put in place to verify the dataset and their contents, and give them a "seal of approval" that they are ethical and trustworthy.
This is why it is absolutely necessary to have a mandatory license for usage in AI training of any copyrighted material, and careful verification of datasets by independent inspectors + total open nature of the datasets for anyone in the public to check their entire contents whenever they want.

Due to the extreme complexity of the datasets required (billions of works, with many tags and metadata associated with each work), it is certainly impossible to guarantee that a dataset is clean and free of non-authorised copyrighted works otherwise

Only then, the AI systems that can be -Proven- to have been trained only on those vetted and approved clean datasets, and nothing else, can at least have a chance to be deemed "trustworthy" ; As already stated many times in other questions.

Anything close to "open source" or non-vetted cannot be trusted, and MUST be assumed by default to contain copyrighted/unclean datasets, knowingly or not.

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

Answer :

I want to add that policymakers should never forget 2 very important and fundamental things when drafting new laws around AI :

first, that any law is meaningless without enforcement.
In addition to the law itself, they should also carefully consider at every step how each of these new provisions passed into law will be enforced, by which agencies, and whether this agency currently has enough financial and personnel means to achieve this new mission or not.

Second, that a for-profit commercial company inherently exists only for profit, and cares only about profit. And thus cannot be trusted to act responsibly and ethically on its own, if there is a greater financial benefit for it to be unethical and irresponsible.
This means that not one single inch of lenience or trust for "self regulation" must be given to big companies involved in making AI systems, or AI Datasets.
Additionally, this means penalties and fines for infringing on AI-related laws must be dissuasive to even multi-billionaire corporations.
If the cost of the fines and penalties are much lower that the potential gain from infringing the AI-related laws, these fines will only be seen by the companies as an "acceptable tradeoff". They will just accept to pay the fine, and walk away with a lot of unjustly earned money only to do it all over again and again.

The best solution to ensure dissuasive but balanced penalties is the same method that is also used by the EU AI Act:
The penalties must be based in % of annual income of the company. (in the case of the AI Act, it can go up to 6%)
If percentages of income are used rather than fixed fines and fees, balance in punishment is ensured : even the richest of multi-billionaire corporations will be dissuaded by those fines, but, a small stratup will not end up being crushed by a fine of multiple billions that is totally disproportionate for them.

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

Answer :

Yes, Clear labeling is absolutely necessary.

It's not just a matter of protecting autenthicity in arts, it's also a basic consumer right : people have the right to choose whether they want to use or view AI generated media or not.
This requirement should always apply in every context : any AI-output media of any kind, anywhere, must be clearly labeled as such without exception. As well as clearly indicate which AI system/model was used.
And mixed works made of a mix of AI generated and creator-made components must all the same clearly indicate what parts of the work are AI-generated and how AI was used.
The companies that commercialize those AI systems are the ones responsible for providing a robust

watermarking/labelling system. If they fail to do so, their AI system will not be approved for public release. The platforms (websites, etc...) on which the AI content is displayed is merely required to use those labels provided by the AI corporation to clearly mark on their interface the AI generated materials.

In case an AI system was allowed for release but it is found after the fact that removal or tempering with the watermarks is ovelry easy, the responsible is the company that made the AI system, for failing their requirement to provide robust identification systems.

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

Answer :

Labeling or watermarking of any kind is easily broken or spoofed by all kinds of malicious actors, making it, unfortunately, probably very difficult to enforce.
Even so, it is still not worthless to pursue. As it will still discourage some malicious users, whereas respectful users will not temper with the labeling, thus allowing it to fulfill its purpose to identify AI generated media as such.

More importantly, intentionally removing the label has at least the benefit of clearly exposing without doubt to everyone the intentions of the malicious actor as malicious, in a legal context, if they still get caugth, they will not be able to play the card of feinting ignorance.

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?*

Answer :

This should be treated as a crime of highly variable severity depending on the case.

For example, a teenager removing a label because they wanted to brag to their friends about being very good at drawing and passed an AI image as something they drew themselves, without trying any sales activities. This would be something fairly triffling and should never be punished severely (although it should still be punished in some form).

On the other extreme end, someone deliberately removing labels to try to impersonate someone else ; or mass posting thousands of AI-generated media to flood media and social spaces with disinformation in order to manipulate public opinion or stock market, or to try to create a political incident... something like this on the other end, should be punished extremely severely with many years of prison.

On the topic of copyright and USCO-related issues, the removal of AI labels to pass AI-generated content as creator-made works for commercial purposes could, in my opinion, be treated as forgery and/or sale of counterfeit goods, and punished according to laws relevant to this (with which I am not familiar).
Or a new law can be written with corresponding punishment specifically dedicated to crime of removing AI labels, if this is more appropriate. Again, I do not have legal knowledge on this.

***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?*

Answer :

There are various AI-based auto-detectors on the market for AI media of various kinds. (text, image, sounds, etc...).
In my personal experience over the past year, and according to what I heard others say too, there is not a single one of these tools that is anywhere near trustworthy enough to draw any conclusions. Some of them "kind of work" in the sense that they get it right more often than wrong, but their error margin is over 15% at best, often more, in most tests I made myself.

My experience, summed up quickly, is that the most reliable way by far to identify AI images is someone who is very, very saavy about artworks and drawing creative process. either because they are a painter/artist themselves, or because they are a very long time amateur that has spent a lot of their free time each day looking at various artworks.

I consistently did better in my tests that the most commonly used AI image detector. Which i found in my testing to make around 15% errors on average (those errors being both false positives and false negatives).

I didn't personally test any other media form beyond images.

***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?*

Answer :

I do not have enough legal knowledge to answer for sure.
But I will argue that this -should- be covered by both general image rights, and by personnal/biometrics data related rights, at minimum, or with even more severe restrictions, if required. If not, then it should.
Using for AI training dataset someone's likeness/voice or any kind of biometric data, etc... without obtaining first their consent of the person through a legally-binding license should be illegal (much in the same way as using copyrighted works without consent).

If someone uses AI systems without permission to reproduce someone's voice, face, or likeness of any kind, and uses this for anything at all without the person's permission, then this should be treated as identity theft/impersonation, and punished with laws according to this.

(On a side note, the same goes with using AI systems to reproduce the style of a particular artist and pass the output as one of the artist's work without their consent : in this case, impersonation/identity theft, counterfeit goods and forgery all should apply.)

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?*

Answer :

There are no such protections currently.
(other than the already existing copyright law, which as explained in question 5, clearly should be applied to AI systems but currently are being essentially ignored by regulating bodies).

I think that this problematic situation can be -mostly- avoided the majority of times anyway if it is made illegal to use copyrighted work for AI training without the copyright holder's consent (see questions 9. and 9.1).
Because, of course, ML AI systems only work as a collage machine that can only arrange together in a patchwork elements taken from the training dataset (see question 7. and 7.1).
So, as long as an AI system does not include any of a specific artist's work in their dataset, it will not be able to replicate this artist's style.

Any and all artists that do not want their style to be replicated by AI systems will simply be free to refuse licensing their work for AI applications of any kind. and not sign any contract that stipulates their work will be used for AI applications (also talked about in question 9.5).

Even in those 2 cases, however, I can still see a situation where all the employers in one specific industry arrange together to only offer contracts that will include using the artists's work for AI. In this case, options for artists to find jobs without having their work used for AI will not exist, thus still forcing artists to sign AI dataset licenses they do not agree to.

In this case, if it truly becomes a problem that reaches this point, then maybe at that time it will be better to just cut it clear and make it so that only the creator of the work and no one else can decide whether the work can be used for AI applications or not.
Not the employer, not the contract stipulations, not even the copyright holder (if a different person than the creator of the work), only the creating artist can decide.