First Name: Camila
Surname: Delgado-Montes

Preface:

I am an artist and a biologist. While I have not published any research with machine learning, part of my studies of GIS (Georeferenced Information Systems) and Remote Sensing color for my research has influenced my understanding of machine learning and its uses, of which it has undeniable useful ones. As a digital visual artist, I also have a perspective based on the use of computer tools, versus the tools taking over the process for you in the creation of artwork. My focus is that of protecting artists from having their livelihoods stolen, but while also allowing researchers to be able to use "AI" (in question marks because while the words Artificial Intelligence are used, I do not believe that the name fits the actual phenomenon/models).

I will not be able to answer all questions, as time is at a premium and some questions are outside of my knowledge and experience. The questions identifying specific types of legal language are beyond my knowledge to answer, for example.

General Questions:

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

I view the risk of the technology as great in scope. Artists are generally already underpaid for their labor, which takes years of practice and effort to master, and I have heard reports of creators being shafted in favor of using "AI" (in quotation marks because the AI that exists today is not an intelligence and does not *choose* and thus does not deserve the title). I have heard reports from other creators – artists – that already usually unstable job security is more precarious are jobs are becoming harder to find due to potential employers starting to use AI to replace them – even though it doesn't have a person actually creating an original expression of an idea, and has many flaws visible/present to the average viewer/public upon viewing the creation for longer than five seconds. I fear that it will be used to further decrease skilled artist's services' market value, who deserve living wages for their work, especially if copyright is given to "AI" "creations". For example, "AI" was

recently used for images in a book in process of creation the popular tabletop RPG game Dungeons & Dragons, and while public outcry resulted in a change in the images used, which were full of obvious flaws, they would have gone ahead with it to try and cut costs and time.

I do see use and benefit in the use of "AI" as a tool for "visual brainstorming", or assembling images that can inspire an artist on the actual, non-"AI" creation of a work – but that would require great transparency on its use, and how the idea and its expression is developed. A hard cut off for inspiration to creation would be difficult to legislate and determine, but perhaps the conditions for "Fair Use" of already existing copyright material would apply here.

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

I do believe new legislation is warranted to protect the copyright of actual artists and creators. It ought to include clear case examples of what fair use is in AI, and what is not fair use – for example, whether training on copyrightable material is allowed without permission from the copyright holder (I would say no), and whether the end product of an "AI" generative process is copyrightable (I would also say no).

Training

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

From my understanding and observation for the visual arts, training materials are "scraped" from the internet, usually from websites containing a great volume of images, without permission of the artists, even by "AI" that is purported to exclude them, due to the mix of opt in or opt out abilities during the timeline of an AI going live, and whether people are properly informed that their material is being used as training material.

For GIS applications, my understanding is that data sets of locations and the remote sensing values (satellite images with different bands of wavelengths) are used for the creation of maps. Most of the time the databases are collected by the very people who did the work of collecting the locations and data associated with them (such as soil samples and nutrient and carbon levels in said soils), which at least avoids out of the problem of using somebody else's work or copyright to create your AI generative product (maps). But some are

used by collating different datasets from different owners, which would require permission from the people providing the datasets.

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)?

I do not know how or how much academic researchers collect images. I am aware of some private companies training their AI of using publicly posted images on google or other areas of the world wide web.

I am of the understanding that academic researchers and private companies using GIS can collect their data from varied sources, sometimes from more purposefully publicly available databases, often in the name of furthering research, sometimes from private sources, such as a worker contracted by a private company to gather the data.

I am not more familiar with the extent of these collections and can provide no further commentary on the subject.

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?

I do not know licensing models. I have seen that many visual images created by artists and thus under their copyright are used to training materials, even without their permission, and that their use is rampant for "artistic" "AI".

From my understanding, copyright is "intended to protect the original expression of an idea of an owner/creator in the form of a creative work, but not the idea itself". By the nature of machine learning, it does not create original expressions – it is built on copying the expressions of others. A prompt creator might have an interesting idea and an original thought as to how to implement it, but they are not expressing it because they are not creating it when they use "AI"; the machine learning is merely copying off existing expressions based on the prompts provided.

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.

My understanding is too poor to really be able to answer this question. I am concerned by the thought that these training images are unable to be truly removed from the training of the AI even if a copyright holder asks to have their image from the database used to train them. Does AI train in different stages? Or do they train from scratch for each iteration? I do not know.

I am also concerned by the examples of AI models reproducing existing images from its training base with just slight variations.

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

I do not know.

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?

As mentioned in 7.1, I do not know, and am concerned about this too.

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?

I have seen examples of near exact reproductions of existing images that were used to train the database. I would suspect that it might be possible to identify a few images used to train based on the similarities.

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.

Images that are released to be royalty and copyright free, made accessible for free for any use, would probably be fair use – but that is not a copyrighted work. I cannot think of an example of a copyrighted work to train AI model used without authorization being fair use. Perhaps if they purposefully collage AI images themselves without using AI to collage it, it would count as fair use –

as in the process of collaging the user would be creating and transforming the work on their own, with their own creativity. Or perhaps, if they do not use the image for earning money, including not using it in monetized youtube videos, etc, could allow a small space for fair use of AI.

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

I do not know enough about the cases to comment on their ruling.

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?

I'm not sure what is being asked – but if an entity collects copyrighted material explicitly for training, then they are part of the copyright infringement of using copyrighted material even if they did not do the training – like an accessory to an act.

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

It does. For profit developers of AI are trying to benefit financially from copyrighted material, which is different from non-commercial or research purposes which seeks to further understanding and knowledge using information but does not commercially benefit from any copyrighted material.

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

I do not know the quantity.

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?

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)?

They should have to affirmatively consent to opt in the use of their works for training materials – at the very least for commercial AI use.

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

I find that at the very least commercial uses require consent. Research is more muddy – I can see an argument for both yes and no in requiring consent. It might depend on the planned use of the AI model being trained.

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?

I know that some art websites (Art Station, Deviantart) are doing Opt Out, and use a "tag" to indicate which art is opted out of the training of their AIs.

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?

I can see that getting consent in advance from all copyright owners would be difficult for all the volume of works needed, but that doesn't stop it from being necessary and right.

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?

I do not know about existing remedies enough to comment. My thoughts are that a fine would be necessary: a fine big enough to make it a good incentive to honor objectives, perhaps based on a percentage of the company's income – a percentage that would impact them. I would argue that whatever remedy exists should be very accessible and easy for the copyright holder, as otherwise it will be practically difficult to protect copyright.

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?

I believe so – an artist's, a human creator's, livelihood, depends on their art being protected. To not protect the human creator from exploitation by AI derails the point of copyright protection, even if the artist was hired to draw/paint for someone else that owns the copyright.

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

I do not know. I'd prefer it if it were official and kept a record, and importantly, for a copyright owner to be able to remove consent later on if they change their mind.

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

I do not know enough about the process of direct voluntary licensing to know the details, but I imagine that while it would be work for commercial AI in creative sectors, it would be feasible – contacting official artists and asking them to fill an online form for voluntary licensing could be done.

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

I do not know

10.3. Should Congress consider establishing a compulsory licensing regime? If so, what should such a regime look like? What activities should the license cover, what works would be subject to the license, and would copyright owners

have the ability to opt out? How should royalty rates and terms be set, allocated, reported and distributed?

I do not know what you mean by this and so cannot comment.

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

I do not know

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

The curator of the training dataset seems like the rational place to require a license.

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

I would expect that it would make commercial art AI more difficult, but seeing as that making it easy would require making it easy to profit off stealing someone's work, I am inclined to consider this a reasonable cost for doing business.

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

 N/A

Transparency & Recordkeeping
15. In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and

disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

Yes. Otherwise, there won't be a way for a copyright owner/artist to determine whether their art was used against their will.

15.1. What level of specificity should be required?

The individual piece itself should be recorded, not just the name of the artist.

15.2. To whom should disclosures be made?

They should be made public and easily accessible, otherwise it will be too hard for an artist to determine if their art was used. If it was possible to use a reverse image search (like google – where you load an image and it searches where that image has been, but in this case, you upload your artwork and the search looks to see whether that art piece was used to train the AI) to identify if an artwork was used, that would make it far easier for objections to be filed if a copyright worker's was unfairly used in a commercial AI.

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

I do not know.

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?

It would make it accessible and possible for actual art to be removed from the training dataset. It would make it possible for objections to be raised in the cases of non-consent of use of an art piece. It would possibly add a cost to the developers of AI models systems, but that seems irrelevant in the face of, again, profiting off stealing someone's work.

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

Preferably, all artists/copyright owners should be notified if their art is used to train an AI model of any kind. Similarly for research use.

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?

I do not know

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

Not for commercial work.

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?

I do not know – I am not a lawyer.

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?

I do not believe AI-generated material needs much in the way of legal protection – the model itself is what requires legal protection, not the product.

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?

I do not know about sui generis right.

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

I do not know enough about the Copyright clause to be able to comment reasonably – I am not well versed enough in legal language to be able to understand it.

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

I do not understand what is meant by implicate the right of reproduction, etc.

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?

Probably demonstrating the access to copyrighted work plus sufficient similarities (such as, if the prompt includes the artist's name, it provides a significant visual similarity to the artist's previous work). Not the style exactly – you cannot copyright style – but if the composition and colors are very similar, that is pretty strong evidence of the use of their work.

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?

I am tempted to say all of them – but most especially, whoever it is that assembled the training materials to include copyrighted material.

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

I do not know.

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

I am not a lawyer – I am unable to answer this question.

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

Labeling or Identification
28. Should the law require AI-generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

Yes – it's important to credit artists, and also to not allow someone to pretend that they are an artist by using AI, as that would be false advertising.

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

I do not know.

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

It should not be usable without permission of that particular person.

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

I do not know.

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

Yes, but only for AI systems, not for humans imitating a style. All artists, all performers, should be protected.

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

I do not know.

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

N/A