General Questions

The Office has several general questions about generative AI in addition to the specific topics listed below. Commenters are encouraged to raise any positions or views that are not elicited by the more detailed questions further below.

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?

a. **Advanced tools used by professionals like the open source software Invoke AI aren't simply entering a single text input to generate an image. They have incorporated these tools into their existing creative process and professional workflows. They might ask the tool to provide them with several versions of a work they had previously created using their own style, take a sketch they've done and generate a rendered image that follows the same compositional structure and outline, or iterate on specific portions of an image that they've created. It's important and critical to accept that these tools will be incorporated into professional workflows, and provide clarity and guidance on how those works will be protected.**

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.

b. **There are and will be significant impacts to creative industries, but any form of licensing regime on the wide scope of content that is used to train AI would be non-viable, and ultimately hinder creatives more than help. AI training is an indirect process of learning patterns between and across datasets, and individual copyrighted works are not infringed in the resulting outputs. As an alternative, I believe artists and creatives should monetize and directly benefit economically from the intellectual property represented by models further trained/specialized on their work, and that there should be protections put in place on the distribution of artist/creative specific models. Artists should have inalienable rights to train models on their own creative work.**

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?

c. **The internet has no borders. We must address the geopolitical reality that inconsistency puts certain countries at an advantage economically. Japan, for example, has already publicly asserted the ability to train models on**

copyrighted content as fair use. Inconsistency here, combined with the inability to ascertain the source and nature of output materials, would put the US at a grave disadvantage.

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.

d. Yes – We need clarity on this emerging intellectual property, and an evolution of laws in order to enable the benefits while curbing harm.
    i. Training should be considered fair use
    ii. Distribution of models merchandised for the purpose of replicating a named artist or the likeness of an individual should be restricted to the named individual.

Training

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

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.

- In the training process, information and patterns about the training data is mathematically determined. For example, after training on a large body of text, the probability of "I was struggling with the problem, but I came up with a great ___" being followed by "idea" is emergently determined to be higher than "pancake". This mathematical property/relationship is represented in the model's weights.

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

- Mathematical patterns are stored as weights/numbers in the AI model. An AI model is a series of numbers that are mathematically influenced during the training process.

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?

- **There are some advances in the area of 'forgetting' concepts, but because it is not memorizing or storing a copy of any individual piece of training material, it is not possible to "unlearn" a particular piece of training material. If a model learned what sunflowers look like by training on 1000 sunflowers, no individual piece of training is fully represented**

in the resulting model. Instead, the concept of a 'sunflower' and its relationship to other elements in the image are the features that are retained in the model.

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?

- Not with certainty, and it would be a near impossible task to determine – Again, because the models are not retaining the training data in its weights.

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.

- In almost every case, except in a situation where an individual is intentionally training on only a single or small set of works and "overfitting" the model such that the only mathematical weights derived are of a single work.

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?

- In Warhol v Goldsmith, the resulting work was clearly related to the original. The relationship between the two was self-evident. Unlike in that case, the purpose and character of copyrighted works is not clearly self-evident in the resulting AI model. Fine-tuning likely has a higher likelihood of being trained on a smaller set of work to specialize the model. There is a more significant risk of a creator's work being targeted for replication in the case of 'fine-tuning', relative to the broad set of work used in pre-training a foundational model.

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?

- It is important to distinguish that any entity which collects and aggregates material is providing a training dataset, by the very nature of information being disseminated. There may be terms and conditions on its use as imposed by the entity, however any distribution of information could be utilized in training in any practical scenario.

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?

- Yes – If these datasets are utilized in the generation of AI models, those models should be made available with open licenses as a benefit to all. This balances the progress made on behalf of society and provides the ability for individual creators to use these foundational models to train for their own uses with their own work and intellectual property.

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?

- Foundational models are trained on vast amount of data. The volume of data used is not being compressed into the resulting model (it would be impossible, given the resulting size of these models). This has a material weight on the fair use analysis.

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?

- **The impact of this technology is broad. It will reshape our society. However, that is an analysis of the technology itself, not any individual model. As a thought experiment, one must simply consider that there are already multiple models with varied sources of data. Any single one of these, regardless of the dataset, have the same competitive consequence – The fourth factor of fair use is clearly intended to focus on works which are specifically targeting the market of the copyrighted works in question.**

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)?

- The practical course for copyright owners is to affix explicit terms & conditions, robots.txt instructions, etc. that restrict use of their work in training, and to pursue violation of those terms as a violation of contract law, rather than a violation of copyright infringement.

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? (48)

- I don't believe this is a viable strategy.

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?

- It is not practical.

9.4. If an objection is not honored, what remedies should be available? Are existing remedies for infringement appropriate or should there be a separate cause of action?

- If an objection is not honored, it will be extremely difficult to ascertain and prove.

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?

- No, as training should be considered fair use. However, a human creator should *always* maintain the right to train an AI model on their own work. This allows for human creatives to maintain the generative capacity for the full body of work they have created over their careers, even if they have competitive pressures in the labor market to sell copyright/license their work to others.

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

- My stance here should be clear, but I believe the only practical course is for this type of training to be considered fair use, and to focus on model distribution that competes (on the basis of violation of trademark or right to publicity)

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

- Licensing for training data is largely infeasible, and poses the same exploitative centralization pressure that will end up pushing down the price of licensing works for these types of datasets if it were to be enforced. Instead, individual creators should have the right to monetize and license models trained on their work.

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?

- No.

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?

- No.

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

- No.

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)?

- As has been stated above, imposing a licensing regime for training data is anti-competitive, and ultimately does no favors to the individual creators. Synthetic data which performs at the same level of licensed work will be capable of competing on price such that any market the licensing regime enables will ultimately fall in, and be non-viable long term.

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.

- No, it is very tenuous and imprecise. You may be able to make a hypothetical approximation, but as noted above these models learn the patterns between different instances of a certain concept, such that they are distinct from any one of the training inputs. The hypothetical approximation I allude to is, for example, that creators of sunflower imagery would clearly have a higher degree of influence over a generated image of a sunflower than creators of unrelated imagery – however, no single creator of sunflower imagery in this case would be able to argue for more influence than another, and the complexity increases exponentially the more concepts are included in an image.

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

- It would decimate US based interests.

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

- This technology does not violate copyright – However its impacts on creators are certainly felt. We need to take a more informed approach to regulating its use, but copyright is the wrong vector for doing so.

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, although unrelated to copyright. AI models should have transparency in order to demonstrate that terms & conditions or other laws have not been violated, separate from copyright.

15.1. What level of specificity should be required?

- A logged hash of the files used for training, along with any accompanying metadata/files (like captions).

15.2. To whom should disclosures be made?

- We likely need a new government body focused on this emerging technology.

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

- There should be due diligence on the nature of training data utilized, and likely a third-party attestation/report reviewing the above records.

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 be an enormous compliance expense, similar to the compliance with Sarbanes Oxley.

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

- None, as noted above.

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?

- As stated above, new laws are needed.

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?

- Yes, when there is creative guidance or input in the same creative dimension as the output. For example, a melody created by a human which is input into a generative system and expanded to a full musical track based around the melody would have provided the core guidance for that output. Similarly, an artist who passes in a drawing and has assistance in completing the work is providing human guidance for the creative output – such that the work *would not have happened but for that creative involvement.* This does not carry to text prompts, which are clearly able to be reused simply without creative involvement.

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?

- It would be helpful to be explicit about the dimensions of human authorship, as copyright laws are vague – however, the spirit of the law is clear enough and can be interpreted to support the above.

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?

- Yes. Enterprises are already using this technology, and they need guidance in order to complete effectively.

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

- If different than copyright, only in the duration of protection.

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?

- Yes. Because the work is creative in nature, and moves the arts forward.

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?

- Yes, although not through training. There are specific processes by which an AI can guided to follow the lines, contours, depth, and other characteristics of an image during the inference process. However, this output would be subject to the same criteria for transformative creative works as has already been litigated in the past – In some cases, where there is no clear tie, it would be very difficult to determine the inspiration as the work has been sufficiently transformed so as to no longer be recognizable. In others (e.g., a similar degree of transformation as Warhol v. Goldsmith) the AI-generated aspect of the work is irrelevant. In either case, we have an existing framework that stands independent of copyright.

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?

- Substantial similarity is accurate, and any other standard would decimate the creative industry, as it would need to be applied to all artists and creative works, human or otherwise. An artist's process is inherently inspiration driven.

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?

- Fair use is the only reasonable path forward on training materials – Partly because the practical answer here is largely "they can't"

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?

- The publication of the generated material is the responsibility of the end user of the system.

25.1. Do "open-source" AI models raise unique considerations with respect to infringement based on their outputs? (53)

- The output of an open-source model should be treated no differently than that of any other model. The responsibility lies in the model's use.

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?

- It is irrelevant, because the use of training should be considered fair use.

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?

- No. AI-generated content will be a matter of degree, not of kind. Labeling is an exercise in marketing, and won't have a substantial impact on usage. Instead, we should focus on ensuring the authenticity of claims of human-involvement in the creation of works, as that will be the determination upon which creative value is based.

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

- The practical reality is that we must be on the defense for AI-generated works, not expecting those who would misuse it to follow laws. Requiring ethical use of AI to be labeled creates an environment where unlabeled works are assumed human/authentic – Unethical abuse of the technology will be unlabeled, and receive the assumption of authenticity by society until disputed. The alternative is that human-originated works are certified as authentic, and fraudsters must make an affirmative claim and undertake effort to forge that credential.

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

- These types of requirements would be largely performative, ineffective, and would not realize the goals they set out to accomplish.

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?

- Context probably matters most, which is where legislation is needed.

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?

- They are all ineffective. We must move towards an assumption of AI-generated content and regime of demonstrating human authenticity in order to separate signal from the noise effectively.

Additional Questions About Issues Related to Copyright

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

- The focus should be on the distribution of models trained to generate material that is targeted at a named individual. This should be protected against.

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?

- Yes.

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?

- No. "Style" is not copyrightable or protected. Many artists share stylistic elements, and art is entirely based around "movements". Allowing this type of protection would strangle creative safety and experimentation in the development of new styles. Instead, the focus should be on distribution of models that are <u>merchandised</u> as replicating an artist's style. The right to merchandise a model trained in an artist's style should be the exclusive right of that artist.

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? [(54)] Does this issue require legislative attention in the context of generative AI?

- The amount of audio needed to replicate the voice of an individual is so miniscule as to be trivial in nature. The issue at hand is the misattribution of statements (or lyrics) made by a voice to an individual who did not say them. This will be a major issue, but attempting to claim two voices are "too alike" with copyright is pointless. Other laws already protect these types of claims.

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

- We are at a crucial juncture in technology and in history. The decisions made will have extreme ramifications on society, culture, and the ability for our country to endure the trying times ahead. Regardless of "ought", it is important to recognize what is – We are facing a technological revolution that requires pragmatism and a deft and informed response.

I am available for further inquiry if I can be of assistance. I am a citizen, and an entrepreneur working daily with this technology. I clearly have opinions – I'm open to different perspectives, as well as supporting discussion, brainstorming, and problem solving to tackle these issues.