My name is Stephen Lindholm and these comments are not submitted on behalf of anyone else or any organization.

I have worked as a lawyer for 20 years, and I also have an advanced degree in computer science. I do not work for any company which uses AI as a material part of its product or development. My thoughts in response to the Copyright Office's questions are below.

My overall view is that the basic issues are not new and little regulation is required.

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

I view generative AI as an unmitigated boon. The question alludes to several potential issues which previously arose with earlier inventions, such as photography, musical recording technology, and digital video and editing technology. The result was that while in the near term specific jobs were lost (there are not many radio stations today which continue to employ orchestras for live music) these skills were readily transferable to other jobs (e.g., studio musicians) and more importantly these art forms were opened up to a vastly larger number of people. To use photography as an example, the average person now takes thousands of photos a year, because technology has advanced to lower costs and skill requirements. We don't need to turn back the clock to the days when any person might have his or her portrait taken once in a lifetime, and if that person died before the portrait was done, his or her family would prop up the decedent on a chair to get a portrait before burial. AI has already been a tremendous asset to society and its positive role will only continue to grow.

More to the point, the accessibility of art is not what devalues existing art, but the distribution. The number of professional musicians has dropped by more than 40% in the US since 1999 and that was not because too many people were able to create their own music. Rather it is because physical music media sales have almost entirely disappeared and the new forms of digital distribution do not pay as much. In the mid 1990s a compact disc would cost about $17 which would be about double that price in 2023 dollars. Nobody is buying CDs for $35 today and that, along with onerous recording contracts and venue contracts, is why professional musicians do not make as much money, which in turn is why there are fewer professional musicians.

No matter how good AI-generator technology gets, there will always be a pinnacle of achievement which will earn a lot of money for the most skilled humans using it. Wedding photographers still make thousands of dollars per event even though, in theory, anyone could take wedding photos on a cell phone or cheap digital camera.

2. *Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?*

No. The only issue I see with AI generated output is that spam is a much bigger problem, e.g., magazines taking story submissions are now overwhelmed with AI-generated trash, Amazon books-on-demand are now filled with AI-generated trash, the world wide web and search engine results are filled with AI-generated trash, etc.

For better or worse this is not a problem that can be solved legislatively as the technology is now widely available (the "horse left the barn" years ago) and so this is not an issue the Copyright Office should address.

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.*

The notion of remunerating copyright owners presupposes that the concept of a "derivative work" should be changed to accomodate the economic interests of existing stakeholders and I see no principled reason why that should be the case. The limits of a "derivative work" is set forth in the Berne Convention and does not depend on a given work is the product of a human brain trained on hundreds of existing works or an artificial brain. There is no reason why, for example, a comic artist being inspired by the style of a particular artist is permissible but an AI brain being trained on that style constitutes an infringement for which the original artist should be remunerated.

This paper (accepted at Neural Information Processing Systems) helps to explain that input data are so thoroughly digested in the process of machine learning that, in a well-trained model, it is utterly meaningless to ascribe any particular part of the artificial brain to any piece of training data: "Towards Understanding Grokking: An Effective Theory of Representation Learning," Ziming Liu et al. https://arxiv.org/abs/2205.10343

Thus, an artificial brain is not a derivative work and therefore any discussion of "fair use" or extracting economic rent (as the question put it, "how licensing regimes do or could operate to remunerate copyright owners and/or creators for the use of their works") is not well-founded under the law.

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?How important a factor is international consistency in this area across borders?*

The People's Republic of China's approach is terrible. In June, the PRC issued its first regulations on generative AI technology, introducing significant obligations for service providers, including content monitoring, marking and data sourcing, while emphasizing the protection of users' personal information through agreements outlining responsibilities.

International consistency should not be the reason for regulating AI more heavily. If another country foolishly chooses to impose limitations on generative AI, the US gains nothing by following.

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

Categorically no. As stated, the copyright liability of a work, under present law, does not depend on whether a human created it based on a life of exposure to existing works or whether an artificial brain was trained on those existing works.

More to the point this is well within the capability of the court system to address just as the courts have addressed previous issues (e.g., whether "look and feel" of a computer program can be copyrighted in *Lotus v. Borland*, file-sharing in *A&M Records v. Napster*, the *Oracle v. Google* fair use case, etc.).

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

Currently any respository of information that can be fed into a machine is in fact fed into the machine, as was seen when early generative AI systems would sometimes reproduce watermarks from stock photo databases.

If the law were changed to require a grant of rights by the owners of the training material, this would favor monopolists such as Adobe, which has set their terms of service such that users' files stored on Adobe's cloud service are fed into Adobe's machine learning algorithm. (Their license states, *"Our automated systems may analyze your Content … using techniques such as machine learning in order to improve our Services and Software and the user experience."*) This would reinforce the dominance of existing monopolists to have the power to impose these terms on millions of users, not to mention reducing the quality of AI models by dramatically shrinking the works available for training. More generally the idea that copyright ownership implies a right of refusal for works to be used in AI training is contrary to millennia of learning, implying essentially that the person who writes a book (or paints a painting) has the right to determine who can read it (or view it) and be inspired to create something new.

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)?*

The dirty secret is that large companies use web-scraped data but do not do the scraping themselves, as that would potentially imply that they are subject to website terms of use contractual limitations even where copyright does not apply. This was seen in the litigation against Google brought by Rap Genius, where Rap Genius demonstrated that Google was using lyrics scraped from Rap Genius's website. Google said, "We license lyrics on Google Search from third parties, and we do not crawl or scrape websites to source lyrics." What Google did, obviously, was to pay a small third party to provide the scraped lyrics. This is essentially a standard practice when one company wishes to use webscraped data contrary to the terms of service of the company hosting the data.

This of course raises an issue of copyright infringement by the scraping company, which is why these companies tend to be small and lightly capitalized. I don't think there is anything the Copyright Office can do here, nor should it.

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?*

This is rarely done unless required for practical reasons (other than datasets intentionally based off of open source or creative commons licenses), because the prevailing view is that a license is not legally required. Adobe's use of its own customers' data is a notable exception however I would hardly call this a great win for reasons described above.

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?*

This is done sometimes. For example, UN and EU transcripts are used to train translation models because they provide a vast quantity of professionally translated text in dozens of languages. However there are limitations, for example this language tends to be formal and therefore less useful for training an AI model to translate social media posts.

It should be noted that non-copyrighted material essentially consists of very old material (due to the continual extension of copyright laws) and government-produced materials from countries where the governmental works are by law in the public domain. Altogether, non-copyrighted material comprises a very limited selection for training an artificial brain. More common is to use open source or creative commons materials, which are those where the holder of the copyright has attached a license with liberal usage rights. The reason that open source or creative commons materials are more common than contemporary public domain works is that (a) historically there was some doubt as to whether copyright could be irrevocably abandoned in all countries and (b) many creators wish to impose some limited obligations on how their works are used.

4. *Are some or all training materials retained by developers of AI models after training is complete, and for what purpose(s)? Please describe any relevant storage and retention practices.*

This I don't believe it is relevant. If someone has access to, say, a stock photos website, and he or she downloads the watermarked samples from the website for inspiration, the mere fact that the samples continue to exist on a computer does not implicate copyright in any way because copyright is concerned when copies are made (or a performance is made). This question is steering close to the misconception that copyright is intended to regulate use of a work, i.e., the misconception that copyright starts from granting all economic rights and all control to the rights holder.

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:*

    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.*

The source materials are digested at the outset of the learning process and those digested materials are used to adjust numeric weights in an artificial brain. The weights represent an "average" from comprehending a vast set of training data. For example, the popular Stable Diffusion generative AI was trained on 2.3 billion English captioned images, as well as 170 million images from another data set. The Stable Diffusion artificial brain can entirely fit into and operate from a single, consumer-grade, eight-gigabyte graphics processor. This is an absolutely massive reduction in size, reflecting how thoroughly the data was digested and "averaged"; if it's not clear how little eight gigabytes is, the largest original iPhone had 16 gigabytes of storage and did not hold 2.3 billion images.

I pause for a moment to note that incidental copies are made in the training process. For example, a photograph may be downloaded from the web, then transferred to a master server, then transferred to individual servers performing the training. Each of those transfers may itself involve several copies; for example, the networking hardware in a computer may receive the file, which is then copied into main memory, which is then copied to the hard disk, and along the way it may be copied into and out of the main processor several times. I think it is not reasonable to apply copyright law to these individual copies, as the photograph was made available on the web in the first instance to be downloaded, and it is the very nature of data that it cannot be used without being copied. This should not be grounds for expansively interpreting copyright law.

However that is not to say that companies should be able to collect training data from around the web and then sell those collections to a third party, because that act of copying is clearly an act of direct infringement in the absence of a license. An analogy may help make my point.

Suppose someone reads an article from the New York Times and summarizes the article in a sentence on his blog, using his own words. This would clearly not be a derivative work, and therefore not infringing, as copyright does not protect ideas only the expression of ideas. (Note, by the way, that this single sentence might represent a 50:1 ratio of "digesting" which is far cruder than what the AI training process does.) However, if someone were to collect all the New York Times articles from the last ten years into a database and sell that database to AI companies for training purposes that would clearly be infringement. This is hardly an issue unique to AI and copyright law does not need to be adjusted in any way.

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

As noted above, the inferences are stored as numeric weights which effectively function as an average of a vast number of input data files.

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?*

It is rarely the case (and undesirable) for an artificial brain to base aspects of its knowledge on a single input data file as this makes the artificial brain useless. Imagine, for example, that you were training a human hardware store clerk and you gave him a single 3" long hex-headed screw and said, "This is steel." If he had no other knowledge of what "steel" is, he would not be able to help any customers who asked for "steel" products because he would not know whether steel was the color, material, length, the height, the shape, the function, the weight, etc. Exactly the same problem arises with training an artificial brain.

As such it is no more possible for an artificial brain to forget what it learned from a single source than for a human to do so. Knowledge by its nature is an "average" from multiple input sources.

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?*

Early in the history of generative AI there were instances of "overtraining" where a generative AI accidentally produced a new work which was a close replication of one of the input works, however there are now means to ensure that this does not happen. As a general matter it is not possible to determine whether one particular work was fed into the training process.

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.*

Fair use is a defense which presupposes that a copyright violation would have otherwise occurred. It is inappropriate to jump to a "fair use" analysis without first determining that the training of AI models itself constitutes copyright infringement.

There are two aspects of generative AI which are subject to copyright analysis: the works created and the artificial brain itself. As for the works created, this question should be left entirely to the courts because it is such a fact-intensive inquiry. In my view, for these works to be found to infringe the training material, the artificial brain itself would also need to be found to infringe unless the artificial brain were given very specific instructions by the end user.

Regarding the artificial brain itself, the mere fact that the numeric weights in the artificial brain are adjusted by processing training material does not mean that the numeric weights constitute a derivative work. The artificial brain has learned typical characteristics of various concepts but

without specifically memorizing any one piece of training material,[1] and the fact that specific aspects of training material may occur in the output when using the artificial brain does not mean that the artificial brain infringes. Consider an artificial brain which has been trained to recognize images of dogs. It is shown thousands of images of dogs, and thousands of images of animals which are not dogs. After training, it has an “average” sense of what is a dog and what is not a dog, which are embedded in its numeric weights. This is similar to the copyright concept of scène à faire, where a case in 1949 held that two works could both represent then-life in the south Bronx using "drinks, prostitutes, vermin and derelict cars.” Similarly, TV and film composers have long written music to resemble well-known music, with features in common different enough to not owe royalties, and painters and illustrators have copied the styles of most well-established artists, and these are not copyright infringements.

I think the most likely scenario for finding copyright infringement is where an artificial brain (by supposition, the brain itself not being an infringing work) is prompted based on a specific work, for example when ChatGPT is asked to summarize a document. The summary of that specific work could constitute infringement, but this question is not AI dependent. Microsoft Word has had an “autosummarize” feature for decades and it too could create an infringing work. Or simply copying the introductory paragraphs from a document would also create an infringing work. In any case this scenario is not actually about AI and it is inappropriate for the Copyright Office to create AI-specific rules based on this scenario.

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?*

These cases are not on point. In *Google v. Oracle*, Google had literally copied 11,000 lines of Oracle’s Java work and distributed them verbatim as part of its Java-like product. In *Andy Warhol Foundation v. Goldsmith*, the Warhol print was almost an exact replica of the original photograph (changed to two colors apparently through a photographic process, then silk-screened) and copies of this near-exact replica were sold for a significant amount of money.

As I have argued above, when someone has made available a work they don’t have the right to control how that work is used as long as copyright is not implicated. Imagine if JK Rowling had published Harry Potter with a big caption at the top of the first page, “You are not allowed to be inspired by this story to write your own young-adult literature featuring a boy with secret wizard powers.” Such a statement would be utterly meaningless not to mention ridiculous. The mere fact that a computer processes the training material rather than a human brain does not implicate copyright differently.

I’m not saying that the collection of training data should be a Wild West free-for-all — most copyright holders of digital media have technological measures in place to prevent the wholesale downloading of their databases. There are many good reasons (such as to avoid load on the system and to prevent piracy) that would also effectuate a copyright holder’s desire to exclude their materials from the training of generative AI, or to be specially paid for allowing their materials to be used in training.

---

[1] It is in fact undesirable for an AI brain to memorize training data exactly; this is called “overtraining” and causes it to perform poorly in the real world because it will only respond correctly to the training data and not real-world data. Thus by its nature, a well-designed AI brain will not constitute a derivative work.

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

As I stated above with my hypothetical about a company that collected New York Times articles and sold them as a bundle, this is clearly a violation of existing copyright law and no change is needed.

There are organizations which have collected open-source and creative commons materials into corpuses for use in training artificial brains, however as these are license-compliant no action is needed by the Copyright Office.

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?*

Fair use is not the correct analysis for reasons stated above (it presupposes infringement), and given that fair use is by its nature a fact-intensive inquiry it is not appropriate for the Copyright Office to weigh in on the topic.

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?*

They use as much as they can get their hands on, billions of data files, as explained above. Inasmuch as any one data file has a miniscule effect on the finished artificial brain, if the Copyright Office insists on pursuing a fair use inquiry, then the larger the number of files used (from the world at large), the stronger the fair use defense against any particular author.

5. *Under the fourth factor of the fair use analysis, how should the effect on the potential market for or value of a copyrighted work used to train an AI model be measured? Should the inquiry be whether the outputs of the AI system incorporating the model compete with a particular copyrighted work, the body of works of the same author, or the market for that general class of works?*

This is precisely the problem with the fair use analysis: That it leads to weighing the economic impact on existing rightsholders without first demonstrating that infringement has occurred. It turns what should be a defense into an affirmative weapon against the on-going advancement of technology.

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)?*

A thousand times no, unless the Copyright Office also believes that books can impose limitations on how readers can be inspired.

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

Nothing in the Copyright Act makes such a distinction, other than that classroom use of copyrighted material is specifically allowed. This question is unmoored from the text of the Copyright Act and smacks of rent-seeking: taxing only those situations where there is gold to be found.

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?*

If the copyright holder does not want his or her work used in this way, he or she should control the distribution of the work appropriately. For example a website can be rate-limited to prevent automatic scraping, and this is alrady commonly done, e.g., to prevent piracy.

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?*

The world is a very large place and if the US attempts to prevent technological advancement from occurring, the development will simply occur somewhere else. Further, the technology will continue to be used within the US, which means the US would shift from being a leader of the technology to being a mere follower.

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?*

This question is so far removed from any current notion of copyright law that it is not meaningful to answer.

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?*

My mind boggles at the notion that this would be the reason that the US decides to import the ill-considered, useless European concept of moral rights.

If an artificial brain is used to create works which defraud or libel someone (including a creator of training materials), of course there are remedies under the law but that does not mean that a creator should have an independent right to object to his work being studied for a particular purpose, including in the training of an artificial brain.

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

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

    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?*

    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?*

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

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

I categorically object to these proposals for reasons stated above.

Further, such a licensing regime would simply not work. The music industry has a blanket licensing scheme because it was built from the dawn of recorded music to enable radio and public performances. Radio and public performances are core to the economics of the music industry, providing income to songwriters as well as radio stations, bars, restaurants and other places which performed recorded music in exchange for advertisers or paying customers. There is no serious songwriter who is not registered to receive royalties, and the blanket license scheme for music dates back practically to the dawn of radio; ASCAP was founded in 1914.

None of this is true for other creative sectors, with respect to royalties for training AI. If a new blanket licensing organization was founded for text and images, there would be more than a century of existing work which was not included in the scheme. Or, if the scheme was made involuntary, the authors of such works would not be identifiable and would not be registered to receive royalties, eliminating any putative objective of the licensing scheme. Likely as well, authors of new works would not bother with registration; unlike songs on the radio where a hit can make a songwriter millions of dollars, no one piece of training data has much value. Would someone bother to register if he might receive a nickel in AI training royalties? I think not.

More to the point such a licensing scheme is not necessary. Collections of large numbers of works useful for training already exist (e.g., the Wikipedia, stock photo sites, Reddit.com) and they already are willing to work with creators of artificial brains. If they impose a fee for downloading vast quantities of data, the fee will also pay for rights to use the data. The blanket licensing scheme would principally apply to images and text scraped off the world wide web but these are precisely the works where automatically identifying authorship on a large scale is impossible.

*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)?*

Even if licensing were necessary to train an artificial brain, this would not present a novel copyright issue. In a chain of licensors, each must have good license to pass on that license and if he or she does not, he or she is liable for infringement. No new action is needed by the Copyright Office.

*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.*

There are undesirable situations of “overtraining” where an AI is too tightly modeled to the training materials. In those situations, the AI is unable to use its learning in new situations and only “knows” the training data. Thus an overtrained generative AI will sometimes fail in producing new output works and will instead imitate specific works used in training.

In those specific instances it is possible to tell that a particular work has contributed to a high degree in the output. However as “overtraining” is undesirable, methods have been developed to eliminate this phenomenon and so, as a practical matter, it is not “possible or feasible to identify the degree to which a particular work contributes to a particular output.”

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

The development of the best-funded and most sophisticated generative AI systems would slow if not cease, in the US, as the return on investment would not be there. The head of OpenAI

(the company with the best artificial brain in the world) said that the cost of developing GPT4 was over $100 million. If this cost were multiplied by the need to pay royalties on literally billions of training documents it would be utterly infeasible to create a new and better artificial brain.

Work would continue inside the US but at a much less sophisticated level, using training materials for which license fees are not due. In other countries, work would continue (without paying these royalties) and the US would lose its edge in AI.

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

N/A

*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?*

Any well-advised creator of a training dataset (intended for licensing to third parties) would retain records of the individual licenses to defend against any future copyright infringement allegation. This is no different from, say, LexisNexis or a film production company storing records of the licenses they've obtained.

1. *What level of specificity should be required?*
2. *To whom should disclosures be made?*
3. *What obligations, if any, should be placed on developers of AI systems that incorporate models from third parties?*
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?*

An affirmative requirement for recordkeeping is ridiculous; if the entity is engaging in activities regulated by the Copyright Act, as explained above, they already keep these records.

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

This question is as silly as asking, "What obligations, if any, should there be to notify copyright owners that their works have been used to build a search engine?"

If someone does not want his works being used in a way he or she finds objectionable he or she should not put them out in the public for anyone to download and view. For example, Matthew Barney does not release his films on home video; they can only be viewed in a movie theater and thus they are not likely to be used to train an artificial brain.

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

If this were true, it would certainly seem to be outside the ambit of the Copyright Office.

*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?*

Most certainly yes. The Copyright Office's efforts to draw a line in the sand around what can be copyrightable (e.g., the "monkey selfie") founders once human intellect and creativity is brought into the issue. With non-AI technologies, virtually any amount of human involvement is sufficient to create a copyright in a work. For example, someone could be blackout drunk, wave around a camera and take snapshots just to be silly ("Look at me! I'm a fashion photographer!"), and that person would have a copyright in those snapshots.

Copyright can arise with instructions not much different from those given to an artificial brain. While casual users of an artificial brain are likely to give trivial instructions just for novelty (e.g., "show me Hodor dressed as Santa Claus"), serious users will give instructions hundreds of words long to precisely state what is wanted. This is no different from a photographer ordering a stage to be decorated and lit a certain way, before pushing a button to snap a photo, where he would be the copyright holder absent other facts (such as a work-for-hire contract as is common in the motion picture industry, or a producer being the "dominant author," see *16 Casa Duse LLC v. Merkin*).

For what it's worth, I think the "dominant author" framework is capable of justifying the human author holding copyright in a work produced by an artificial brain. A human might wake up one day and see a sunrise and be inspired to paint a painting, whereas an artificial brain will never make art out of genuine inspiration. An artificial brain is entirely subservient to the human's will and therefore it is the human mind which is the dominant author. This is similar to how a camera is subject to a human's will, no matter how slight, and therefore photographs taken have a human's copyright, except in the very rare cases of, e.g., a "monkey selfie."

The category of computer art has been well-established for decades (see, e.g., the recent retrospective at the LA County Museum of Art). Someone could devise an equation which produces a beautiful piece of art when computed a million times; indisputably that person would have copyright in the final work. The equation could even incorporate an element of randomness. This is not different from someone commanding an AI model to produce art according to specific instructions.

And lastly, the fact that the work of art was not fully conceptualized by the human at the time he gave direction to the artificial brain (e.g., he might have asked for a "fairy" but not known exactly what the fairy will look like) does not, under existing practice, defeat copyright either. Plenty of artists work in process-directed mediums, for example, by placing gunpowder on a canvas and igniting it to produce an interesting pattern, or potters who make raku pots the coloration of which is largely outside human control.

There is effectively no hard line that can be drawn against the use of an artificial brain as a tool for art that does not also stomp over decades (if not centuries) of established art in other media. The copyright office's desire to limit the scope of copyright to human expression of ideas must necessarily stop at the "monkey selfie" to maintain logical consistency.

*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 seems unnecessary to state whether something produced by AI under human instructions has copyright in it. The first court case to test the issue will most certainly have strong facts in the human's factor and thus the court will find that the human does have a copyright, which means the Copyright Office need not make definitive pronouncements.

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 don't see what the computer code has to do with the issue. The works produced by an artificial brain are much more the product of the weights in the artificial brain, together with the human instructions, rather than the small amounts of computer code which operate the artificial brain. To assert that copyright in the computer code transfers to the works produced would be like asserting that Canon holds the copyright in all photos taken, because they designed the lenses which cause the film to record a specific representation of an image. Copyright is not a but-for test; it's based on who expressed an idea.

Further, I don't think the copyright in the weights of the artificial brain transfers to the works produced either. The artificial brain requires instruction at the time of use in order to know what do to. This is just like how a musical instrument requires human direction to produce music. We don't say that Steinway gets the copyright for any music composed on a Steinway piano, because the human brain is where the creative control lies and the piano is just a tool acting under the human brain's command.

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?*

We certainly don't need any more sui generis "intellectual property" to be created in the absence of a demonstrated problem.

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 think the answer is evidently "yes" based on arguments I made above. The question could be altered to say "Does the Copyright Clause permit copyright protection for photographs? Gunpowder art?" Obviously the answer is yes to those, and as with them "AI-generated material" is always done at the direction of a human.

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?*

Sure, just like someone could use a camera and props to replicate existing artworks, someone could use an AI generator with very specific instructions to replicate existing artworks. There would of course be an issue of "fair use" as artists always have a bit of artistic license (compare *Mattel v. Walking Mountain Productions*, which found a right to use photos of Barbie dolls as an art piece, with *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*) but the analysis would start with a finding that the replica is a derivative work.

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?*

I have a hard time understanding why it (or other tests such as direct copying) would not be adequate.

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?*

First, it is never necessary to prove literal access to the copyrighted work; circumstantial evidence is sufficient.

Second, this sort of issue arises almost exclusively in the context of music copyright infringement suits because these cases are so weak to start with. It's often the situation that similar music is created independently by different musicians because there are only so many chords to start with (and fewer still which have the cultural associations making them suitable for pop music) and the chord progressions are dictated by music theory. In successfully defending the recent lawsuit against him, Ed Sheeran testified (and demonstrated with his guitar) as to 101 examples of other well-known songs using the same chord progression as the song he was alleged to have infringed, and he stated that these examples were just "scratching the surface." Thus, in order to bring a lawsuit for infringement, demonstrating access to the work is essential.

Further, these music copyright lawsuits are inevitably brought by greedy heirs (e.g., the suits against Ed Sheeran, Led Zeppelin, Pharrell Williams etc.), not the original artists whose work is alleged to have been infringed. I frankly do not think we need to grease the skids for these frivolous cases by requiring documentation to be kept.

*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?*

This is really a fact-dependent inquiry that would depend on how the artificial brain was used. If someone gave generic instructions but the artificial brain replicated one of the images in its training set, this is clearly the fault of the creators of the artificial brain (as "overtraining," described above) and they should have taken measures to prevent it.

If someone gave precise instructions which caused the artificial brain to replicate an existing work, then the fault lies with the user.

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

I don't understand what is being alluded to here. "Open source" is not a new type of copyright law, it is merely one type of license and the same principles apply as with any other copyright license.

*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 does not, because despite how 1202(b) appears to read, it does not cause liability for the *"universal possibility of encouraging infringement."* (*Stevens v. Corelogic*) Instead, violation of this part of the DMCA requires knowledge that this will cause identifiable infringements in the future. *"[A] plaintiff bringing a Section 1202(b) claim must make an affirmative showing, such as by demonstrating a past "pattern of conduct" or "modus operandi", that the defendant was aware or had reasonable grounds to be aware of the probable future impact of its actions."*

As training an AI is not an infringement, *ipso facto* there is no violation of 1202(b).

Additionally, the work is reduced so much in the training process (see above, referring to the billions of images used to train an artificial brain that can fit into and operate from a consumer-grade graphics processor) that the AI model cannot be said to be the original work, and thus it is not possible for a claim to arise based on having the original work detached from the CMI.

*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

*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?*

I have four words for this: "Good luck with that."

There is no practical way to require this, such a requirement would be like requiring consumers to keep the tags on their mattresses.

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

Maybe the person who snoops in trash bins and rats out the people who don't rinse their yogurt containers? It would have to be someone with a lot of free time.

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

Yes, because several artificial brains are freely available, and any part which serves the function of labeling or identifying the output can be easily removed.

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?*

This seems not much different from requiring all Photoshopped images to be labeled and would be swiftly defeated under a First Amendment challenge.

*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?*

The leader in artificial intelligence is OpenAI and the way they have described the situation, such efforts look hopeless. https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text *"In our evaluations on a 'challenge set' of English texts, our classifier correctly identifies 26% of AI-written text (true positives) as 'likely AI-written,' while incorrectly labeling human-written text as AI-written 9% of the time (false positives)."*

The limitation is that these tools don't work. It's elsewhere been stated, with respect to bear-proof containers at Yosemite National Park, "There is considerable overlap between the intelligence of the smartest bears and the dumbest tourists." It's not surprising that generative AI is smarter yet.

*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?*

There is nothing in these statutes, to my knowledge, that depends on how a work was created and I don't see a novel issue here.

*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?*

God no, we have enough sui generis intellectual property laws as it is and the states have not been hesitant to expand personality laws whenever the courts have ruled in favor of free speech (see *Lugosi v. Universal Pictures* and the laws passed thereafter, and the new personality laws in California and New York).

*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?*

Only if human artists are also forbidden from copying the style of an artist; there is no reason that the way a work is created should affect its copyright status.

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

Section 114(b) of the Copyright Act is terrific and should not be changed.

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

N/A