Dr. Jennifer Pullen's comments on copyright and AI

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)?

Yes, copyright owners should have to explicitly consent to have their works used to train AI. Any use of copyrighted material for training AI without explicit and compensated consent is in effect theft. I am a college professor, and if a student uses other written works to write a paper and doesn't cite them, it is plagiarism. Writing a paper using language and ideas of others rearranged and partially in one's own words is a form of plagiarism called patch-writing. When it comes to written texts, AI generated fiction, for example, is patch-writing. If AI programs were human students, AI generated text would be sufficient for them to fail an assignment, or even a class, because of patch-writing and a lack of acknowledgement of sources.

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

10. If copyright owners' consent is required to train generative AI models, how can or should licenses be obtained?
Those seeking to use the copyrighted work should have to contact the copyright holder directly and obtain their consent, as well as compensate them. If consent is refused, or compensation is declared inadequate by the copyright holder, then any use of copyrighted material should constitute intellectual property theft and be punishable to the fullest extent of the law.

10.1. Is direct voluntary licensing feasible in some or all creative sectors?
Yes, most artists have websites or social media where they can be easily contacted. Those who, for reasons of privacy prefer all inquiries to be directed to their agents, agent contact information is easily available. For example, I am a college professor and a fiction writer. I can easily be reached via social media, my university email and phone number. However, if I were a famous writer whose contact info was kept confidential for that reason, my literary agent's contact info would be accessible, and he would be the appropriate person to contact about licensing my works, and he would be able to contact me. The other arts work similarly. Companies that use generative AI or create generative AI should have as much responsibility to get permission and compensate copyright holders as individuals. For example, I recently wrote a textbook for Bloomsbury Academic. As the writer, I had to contact authors to use their works in an anthology, get written permission, and arrange compensation. Any use of someone else's work, even if I simply referred to it, requires citation and acknowledgement.

10.5. Should licensing regimes vary based on the type of work at issue?
Yes, not all artistic mediums are equivalent. Additionally, not all artistic mediums have collective bargaining bodies that would make any kind of mass licensing possible. Screenwriters have a union, for example, fiction writers do not.

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, all such information should be required by law to be open, searchable, and easy to access.

15.1. What level of specificity should be required?

Copyright holder name, and name of their work that was used.

15.2. To whom should disclosures be made?
Publicly available, and disclosed to the US copyright office.

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 doesn't matter how much it costs. Artists are much more vulnerable than technical companies, because they are already on average poorly compensated, and so they often lack the funds or resources to enforce compliance with the law, or to protect their copyright. Thus, the burden of disclosure and record keeping should be on the companies using and making the generative AI technology.

16. What obligations, if any, should there be to notify copyright owners that their works have been used to train an AI model?
AI should not be allowed to use copyrighted material without explicit permission of the copyright owner, period. So obviously, permission requires contact and thereby notification.

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?

This is a tricky question. If the AI someone is using was trained on copyrighted work without the consent of the copyright owner, then anyone using it is guilty of plagiarism, so one wouldn't want to be considered the author of a plagiarized work, would one? However, if for example, an author took an AI program, and input only public domain works, or input only their own work, and the AI program wasn't trained on copyright materials beforehand, and then whatever output that AI created, the human author them extensively modified and revised, it could arguably be considered a case of copyright eligible work, that one could ethically claim authorship of.

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?
Copyright protection for the computer code seems sufficient. Any other copyright should involve work that is human authored. We don't want to create a system in which humans are disincentivized to author their own works without AI assistance, or in which works created with AI are treated as equivalent to something written by a person. The precedent as set by the recent screenwriter's strike seems like a good compromise in this regard.

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 should. This is vital for consumer protection purposes, and to protect the rights and economic viability of artists.

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?

I am a writer and a professor. AI generated work has already caused several magazines, such as the award winning science fiction magazine, Lightspeed to have to close early to submissions because of the substantial burden generated by having to sort through an unexpected influx of submissions because people submitted AI generated work. It is an undue burden on an industry that already operates on a shoestring staff. Additionally, authors are running into problems where AI generated copycat works are being self published on platforms like Amazon and elsewhere, and competing with their genuine works. The AI written material is terrible, but it takes time and effort for folks to sort out the AI generated work from the real thing, and in the meanwhile, those without information literacy can be fooled into thinking they are purchasing or reading a human made work.

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

From Copyright.gov:
"Under the *fair use* doctrine of the U.S. copyright statute, it is permissible to use limited portions of a work including quotes, for purposes such as commentary, criticism, news reporting, and scholarly reports. There are no legal rules permitting the use of a specific number of words, a certain number of musical notes, or percentage of a work. Whether a particular use qualifies as fair use depends on all the circumstances."
https://www.copyright.gov/help/faq/faq-fairuse.html#:~:text=There%20are%20no%20legal%20rules,depends%20on%20all%20the%20circumstances.

Thus, no percentage of copyrighted material can be used for commercial purposes without permission. Unless someone who creates a generative AI can guarantee that it won't be used for commercial purposes, then they must get permission. Otherwise, only works in the public domain should be used.