

Drew Smith - Opening Statement for Artificial Intelligence Study Submission

I cannot speak to every question that has been asked as I am not an expert on AI. However, I can speak as a 27 year old artist whose life and career have been severely altered by the popularization of generative imagery AI models. These popular models such as Stable Diffusion and Midjourney are unethical and infringe on the rights of artists and individuals around the globe whose images were scraped and used in the training data without consent.

Without regulation, these generative image models have already taken jobs from human artists. In addition to the career and employment opportunities lost, we are also forced to fight just for our rights to be recognized, as many would like to claim that our copyright is invalid once they feed it into a machine. Were the process behind these models carried out by an individual, it would be easily identified as infringement. The speed, scale, and granularity of what is done by these models helps obfuscate that reality, but still does not change what it is: infringement.

Copyright and intellectual property law has helped foster creativity and innovation by allowing creators to protect our ownership of what we create. It empowers us to choose and control how our work is used, appropriated, derived, and remixed, ensuring we never need to compete with our own creations. With the development of technology advancing faster than the law, we have lost that protection and suffered for it, both financially and mentally.

It's for these reasons that new legislation is incredibly important. Current models have been able to get as far as they have due to the lack of regulation on the emerging technology. They are taking an "ask forgiveness, not permission" approach that is unacceptable and aims to undermine ownership rights by asserting that the ends justify the means, actively harming the people who made their work possible in the first place

Having one's photos, art, or other imagery used in training data should be an OPT-IN process, full stop. It is not reasonable for any individual to have to seek out all instances where their work might be used and request to opt-out, especially when new instances could be created at any time. Models that have been trained on stolen, scraped data should be destroyed, as it is not possible for them to "forget" data retroactively. Additionally, those who consent to their works being used in training datasets should be compensated fairly, and what constitutes "fair compensation" should be determined by artists and those whose work will be used.

I have dedicated the majority of my life to honing my craft and only just entered the workforce a couple years ago. As things currently stand, I'm immeasurably afraid that my career could end before it truly begun. I will not claim that AI does not have its uses, that it cannot be leveraged as an effective tool in many ways, but without regulations that ensure that humans, as well as their rights and creations, are always and forever valued and guaranteed over those of a machine, there is an unavoidable risk of humans being displaced and made to suffer as a result.

Below I will address some of the specific questions where my knowledge and expertise may be applicable:

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?*

See my opening statement.

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?*

Yes, as a graphical artist, absolutely. See my opening statement for more details.

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.*

AI Art and its Impact on Artists: <https://dl.acm.org/doi/10.1145/3600211.3604681>

While not a traditional paper or study, the advocacy group “Create Don’t Scrape” hosts a collection of resources explaining the impact of AI on artists and the unethical practices of common models:

<https://www.createdontscrape.com/pretrainingfine-tuning-why-you-need-to-know>

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

I strongly disagree with copyright exceptions for text and data mining that could permit use of copyrighted material to train AI systems. I would hope that the USA will lead the world in strong protections for artists, writers, and other sources of training data.

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.*

Yes, see my opening statement.

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

Many copyrighted works that were posted online have been scraped independently from online sources like ArtStation or sourced from large databases, such as LAION-5B.

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)?*

Large databases assembled by a third-party, such as LAION-5B, are often used as training data.

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?*

To my knowledge, there are very few, if any, instances of work being licensed for training data. I believe Adobe is doing something with their own model, but I don't know the details personally.

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.*

None in my opinion. I am not familiar enough with case law to provide examples.

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)?*

Opt-in, full stop. Opt-out is inherently problematic and unreasonable to ask of individuals. Issues regarding licensing and

9.1. *Should consent of the copyright owner be required for all uses of copyrighted works to train AI models or only commercial uses? ⁽⁴⁷⁾*

Consent should be required for ALL models, not just commercial uses.

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?*

If the developer of a machine learning model cannot get consent in advance, it should not exist. Inconvenience should not override a creator's rights.

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?*

The artist should be compensated retroactively regardless, and should be able to request the destruction of any model trained without their consent. This may seem extreme, but would de-incentivise those who create models to use unauthorized work.

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?

Yes. One solution would be to dictate these terms in the WFH contract, but this would be difficult to apply retroactively and could result in artists who request these terms to be denied work. This is an issue I feel someone with more knowledge would be more equipped to address.

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

Licenses could be handled individually or based on a standard license for the model. The specifics of said licenses should be developed with artist/creator input prioritized.

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

It might become much more expensive (perhaps prohibitively so) to require licensing, but if a business cannot afford to operate while fairly and legally compensating its contributors, it does not deserve to continue operating.

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, absolutely, to all questions.

15.1. What level of specificity should be required?

At the bare minimum: the source data used, how it was sourced, and the author of the data

15.2. To whom should disclosures be made?

Disclosures should be public and transparent to all. If data is too sensitive to be public, it should be available to copyright holders.

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

Developers should publicly state any third party sources they use in a way that is easily accessible to the public.

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

Developers should make every possible effort to inform copyright holders that their work has been used.

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?*

No. Humans using an AI are not authors of the materials produced unless they are the sole author of the training data and code behind the process, period.

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?*

Yes, this seems reasonable.

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?*

No.

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?*

No.

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 believe a new standard will likely be necessary.

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?

It is very difficult for copyright owners to prove this, as machine learning models have capabilities beyond human ability. Existing rules are likely not enough, hence the need for new standards.

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 developer of the model should be held directly liable. Systems incorporating the model would be secondarily liable.

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

Yes, as we have seen open-source models be used to circumvent existing laws. Open source models should not be able to have their derivatives be commercialized.

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?

Absolutely. In an age where misinformation is running rampant, this is more important than ever.

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

The individual who generated the work as well as any who choose to use/display that work.

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?

Bare minimum, removal of the generated work until it is properly labeled. More serious consequences could be discussed if the lack of label caused harm.

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? 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?

In my opinion, AI-generated material should not have or retain any rights.

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?

All living artists should be protected from AI systems trying to replicate their work. Models that do so are in direct competition with their livelihoods and actively harm their attempts to find work.