Comments on Artificial Intelligence Notice of Inquiry

88 Fed. Reg. 59,942

Submitter: W. Y.

Thank you for taking the time to hear from stakeholders on the issue of copyright as it relates to the current generation of artificial intelligence (AI) and machine learning (ML).

I am an individual with background in both the visual arts and computer science[1]. My background's relevancy is as follows: I am a visual artist wholly empathetic with the concerns other creators have over infringement of copyright, and I am not fully ignorant to the realities of artificial intelligence as far as current technology is concerned. When obtaining my degree in computer science, the mandatory 1-semester ethics course seemed to lack polish (which leads me to believe it was new to the curriculum) and to fall short of engaging. My experience with the course showed me how often obtaining a full understanding in ethics *lags* behind all other aspects of the field—surprising, considering how applied computer science is irrevocably tied up with actual human interactions and international society.

I hope to contribute a *cross-disciplinary* perspective. All views expressed are my own. As primarily a visual artist, my responses will be skewed toward the visual arts. I was motivated to comment by my growing concern over the continued feasibility of the creative arts as a sustainable source of income and the threat to the integrity of copyright (can existing codes of copyright function in the presence of data laundering?)—both of which threaten the advancement of the arts.

---

[1] 3-year BSc in Mathematics and Computer Science from a UK university.

1. 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?

For additional personal background, I have been drawing and painting with both traditional and digital media for over 13 years. I formally studied the visual arts under a 4-year BFA program in Graphic Design from a US college, including a full year of fundamentals (life drawing, design, color theory, etc.) required of all students irrespective of discipline (including photography). The curriculum did not focus solely on traditional media but also encouraged exploration of digital media and modern technology, including the incorporation of microcontrollers (such as Arduino kits) and virtual reality. In this sense, the current generation of AI and ML provides the benefit of an *additional medium* through which artists can express matters such as: society's increasing co-dependence with technology, the sociological and psychological impacts of modern technology, etc. In essence, striking commentary with the social benefit of provoking a change in thought or realization in its viewers.

However, the current technology in question also poses actual realized harm—not the visions of doom often heralded by the news.[2] [3] [4]

Furthermore, generative AI exacerbates existing exploitative issues threatening workers' livelihoods. I think we should be wary of naïve analogies to

---

[2] See a plethora of examples in the article "AI Causes Real Harm. Let's Focus on That over the End-of-Humanity Hype" by Emily M. Bender and Alex Hanna: https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/

[3] See also the entirety of: Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March 1). On the dangers of stochastic parrots: Can language models be too big? 🦜. *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

[4] See the specific harm to the creative arts and artists: Jiang, H. H., & Brown, L., & Cheng, J., & Khan, M., & Gupta, A., & Workman, D., & Hanna, A., & Flowers, J., & Gebru, T. (2023, August 29). AI art and its impact on artists. *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 363–374. https://doi.org/10.1145/3600211.3604681

photography[5] and automation in industries. Generative AI does not merely "replace artists"—rather it fully replaces all aspects of *intentional decision-making*, which is a problem with implications as myriad as avoidance of accountability, in favor of a system that hallucinates meaning from what is functionally white noise (as in diffusion models). The fact that the outputs of such systems manage to be convincing enough (though often fall apart with close scrutiny and a trained eye) is solely through the unauthorized inclusion of billions of training data.

Your office notes there is disagreement on whether the unauthorized inclusion of copyrighted works in training datasets is infringing.[6] For AI models where the outputs resemble the inputs—as is the case in generative AI—my view is that such unauthorized inclusion is in fact inescapably infringing.

Without protections for copyrighted works from unauthorized ingestion into training datasets, users of this technology actively engage in industry-scale automated "style mimicry,"[7] or "style transfer," whose product is sometimes referred to in computer vision and machine learning literature as a "pastiche."[8] I consider style mimicry to encompass attempts at replicating a distinguishable style (as "in the style of"), especially through the explicit use of an exemplar (example in the targeted style) present in the training data or given as an input. Note that successful style mimicry is only possible with the mass ingestion of copyrighted works (often including the targeted style), and ensuing mimics can emerge in the order of hundreds of thousands in a vanishingly short period of

---

[5] Photographers, too, must consider fundamental matters as composition (framing and precise choice in where to aim the camera), lighting, all foreground and background elements, colors, etc. Cameras with a powerful array of features allow for fine-grained control whose behavior is well-defined—far from a black box system.

[6] U.S. Copyright Office. (2023, August 30). *Artificial Intelligence and Copyright.* Retrieved October 23, 2023 from https://www.copyright.gov/ai/docs/Federal-Register-Document-Artificial-Intelligence-and-Copyright-NOI.pdf

[7] For a definition of "style mimicry," see §2.2 of: Shan, S., Cryan, J., Wenger, E., Zheng, H., Hanocka, R., & Zhao, B. Y. (2023, February). Glaze: Protecting artists from style mimicry by text-to-image models. *arXiv*. https://doi.org/10.48550/arXiv.2302.04222

[8] For the definitions of "pastiche" and "style transfer," see the first paragraph of §1 of: Dumoulin, V., Shlens, J., & Kudlur, Manjunath. (2016, October 24). A learned representation for artistic style. *arXiv*. https://doi.org/10.48550/arXiv.1610.07629

time—far outstripping the production rate possible of human artists that mimic established styles.

Outside of the context of ML, the relationship between art and pastiche—defined by the American Heritage Dictionary of the English Language as "[a] dramatic, artistic, literary, or musical piece openly imitating the previous works of other artists, often with satirical intent"[9]—as it relates to plagiarism, ownership, and copyright already drew contention.[10] The text "Appropriation, Homage, and Pastiche: Using Artistic Tradition to Reconsider and Redefine Plagiarism" by Joan A. Mullin explores the blurred lines between plagiarism and homage—even when students of the arts are encouraged to appropriate in the development of their skills—in a series of interviews with higher education faculty across artistic disciplines. Mullin notes that although collaboration fosters a healthy artistic tradition, ethical and copyright issues arise as follows:

> Contracts can take away all artistic rights; ideas can be manipulated just enough so that legal claims can't be made; a young artist may think he owns material, only to find others making profits from it and claiming ownership. What in the past may have been produced collaboratively may now be subject to negotiation because one in a group seeks ownership through copyright. (Mullin, 2009, p. 113)

Applied to the case of generative AI, corporate profits are made off of copyrighted works that are unscrupulously trained on. As ethical qualms already plague the reality of student development in creative industries, the added dimension of generative AI's ability to mass-replicate problematic "works" (which achieve style mimicry of living artists largely *because* of the ingestion of copyrighted works) only exacerbates the issue. All the while, these mass-replicated "works" flood the creative space with for-profit outputs that contribute little to the progress and advancement of the arts.

---

[9] Other definitions seem to omit the "satirical intent" aside. The source of this definition is: The American Heritage Dictionary of the English Language. (n.d.). Pastiche. In *AHDictionary.com dictionary*. Retrieved October 25, 2023 from https://www.ahdictionary.com/word/search.html?q=pastiche

[10] Mullin, J.A. (2009). Appropriation, homage, and pastiche: Using artistic tradition to reconsider and redefine plagiarism. *Who owns this text?: Plagiarism, authorship, and disciplinary cultures*. (pp. 105–128). https://doi.org/10.2307/j.ctt4cgn56.7

Furthermore, I believe industry-scale automated style mimicry harms living creators, and hence the advancement of the arts.[11] These outputs actively compete with living artists and dilute their individual brands.[12] In a competitive industry where students are often told they will be "judged on [their] weakest piece" of their portfolio, the rampant dilution from style mimicry on a massive scale can damage their online presence.[13] With the ability of social media to rapidly propagate information, the first impression a prospective client holds may be predicated on (unauthorized) automated pastiches that wholly misrepresent an artist's body of work, all the while including a (misleading) mention, tag, or other metadata referencing said artist.[14] Even when the artist's name is not paired with the prompted style mimic, popular artists' styles are distinctive enough (to not be confused with similar styles) and recognizable enough (for audiences to recall the artist's name), and the style mimic is considered by general audiences to be close enough that the association is indelibly fixed. This association again leads to dilution of the artist's brand.

To equate industry-scale automated pastiches with "flattering" homages of the past would be naïve. Far from an expression of flattery or "exposure" (payment via exposure is an age-old issue plaguing creative industries), artists have expressed fear that the discoverability of their work will be overwhelmed (on the order of hundreds of thousands for popular artists, and counting) of generative AI mimics.[15] Also worrying for artists concerned with style mimics—and for individuals concerned with privacy—is the nonzero amount of training data "memorization" possible, even in proprietary models, which suggests that

---

[11] The previously referenced paper "Glaze: Protecting artists from style mimicry by text-to-image models" covers many cases of harm to living artists in §1 and §2.2. Also see §3, which includes relevant concerns voiced by artists, as well as survey results. https://doi.org/10.48550/arXiv.2302.04222

[12] See the interview with illustrator Hollie Mengert at: https://waxy.org/2022/11/invasive-diffusion-how-one-unwilling-illustrator-found-herself-turned-into-an-ai-model/

[13] Marques, B. (2020, September 9). The ArtStation guide to going from student to professional artist. *ArtStation Magazine.* https://magazine.artstation.com/wp-content/uploads/2020/09/ArtStation-Guide-to-Going-from-Student-to-Professional-Artist.pdf

[14] See a prior footnote regarding the case of Hollie Mengert.

[15] See the case of digital artist Greg Rutkowski, whose very name is often *suggested* as a prompt for generative AI: https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/

diffusion models are capable of outputting near-exact replicas of images from their training data.[16] ("Memorization" is especially defined in the referenced paper by Carlini, et al., who note the ethical and privacy issues currently posed and identify a likely cause as the duplication present in training datasets.) It seems that the best measure of safety a creator's work has against blatant infringement corresponds to lower representation of their work in the training data—and really, creators would prefer it be *zero* (for unauthorized works included).

Present use of unregulated generative AI models leads to many artists (including myself) considering removing work shared in public online spaces (at the high cost of visibility), and for the remaining shared works: reducing the resolution, obscuring with watermarks, or applying adversarial perturbations (à la *Glaze*) that are visible on higher settings.[17] In short, reducing the quantity and quality of created works that can be publicly experienced, to the unfortunate disservice of the art community and wider public.

Morale is also an issue: many creators feel a reduced desire to create. Very likely, enrollment will lower in creative programs and fewer will enter creative industries—particularly if generative "AI assistance" is pushed in the industries, which could (and likely does) eliminate entry positions, halting newer voices with fewer options of entering the industry.[18] I myself felt forced to abandon my present path as a work-for-hire concept artist in favor of self-developed projects, a pivot I understand not everyone is privileged to make.

Even students relying on AI assistance in an attempt to improve their skills risk entirely bypassing vital processes of learning (developing strong fundamentals, acquiring a sense of why certain things work, trial-and-error, visual artists

---

[16] Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramèr, F., Balle, B., Ippolito, D., & Wallace, E. (2023, Jan 30). Extracting training data from diffusion models. *arXiv*. https://doi.org/10.48550/arXiv.2301.13188

[17] See especially §3.1 of "Glaze: Protecting artists from style mimicry by text-to-image models."

[18] Carter, T. (2023, October 1). Workers are worried about AI taking their jobs. Artists say it's already happening. *Insider.* https://www.businessinsider.com/ai-taking-jobs-fears-artists-say-already-happening-2023-10

studying from life, etc.). In an online chat server for artists honing their skills, I was disturbed to see some beginner artists asking generative AI to generate variations on their initial sketches in an earnest attempt to study how they could be improved upon—the results were variations that only rearranged elements and by my eye did not constitute an improvement. Using generative AI in this way is certainly a misguided endeavor: the models cannot be directed in this manner (they have certainly not "learned" what is an appropriate incremental "improvement" for a student). All this is why I worry about the stagnation of the creative arts.

Currently, some generative AI models attempt to address the matter with an opt-out system.[19] In a response to a later question, I will elaborate further on the weaknesses of this proposed solution. Although opt-out is only a stepping stone to stronger protections for creators, some services provide a path for opting out via Digital Millennium Copyright Act (DMCA) requests.[20] For such requests to be honored, the following must be explicitly enshrined in law: that the unauthorized use of copyrighted materials in training datasets or as (user) inputs be deemed copyright infringement.

With no small measure of hypocrisy, many generative AI models—including proprietary ones—explicitly disallow users from incorporating those same models' outputs into training data for (competing) AI models.[21] Although at face-value this policy appears to be a standard "non-compete" clause, it seems these companies are patently aware of the exploitative nature of their training datasets.

Take for instance the domain of image generation. The non-profit organization LAION has freely, publicly released large datasets for ML research purposes—including the well-known LAION-5B, composed of billions of image-text pairs (in

---

[19] See OpenAI's opt-out form: https://share.hsforms.com/1_OuT5tfFSpic89PqN6r1CQ4sk30

[20] See the DMCA policy of Civitai, a platform that hosts and distributes open-source generative AI models: https://civitai.com/content/dmca

[21] See multiple quoted terms of service here: Barr, A. (2023, June 3). AI hypocrisy: OpenAI, Google and Anthropic won't let their data be used to train other AI models, but they use everyone else's content. *Insider*. https://www.businessinsider.com/openai-google-anthropic-ai-training-models-content-data-use-2023-6

the form of URLs) scraped from the internet, and whose datasets were used to train Stable Diffusion, Google's Imagen,[22] and other *commercial* (as opposed to for research) AI models.[23] Thus, terms forbidding ingestion of commercial AI models' outputs into further training datasets suggest that these technology companies believe the following: that certain images do *not* belong in training datasets, and that the inclusion of certain images in training datasets induces (unfair) competition. This is not unlike the stance creators take regarding the works they create and/or hold the copyright of; incidentally, copyrighted works likely form the bulk of training datasets[24] used in commercial AI models. Therefore, it seems that creators' fears of unfair competition[25] are substantiated by the very terms of service of the highest-valued AI models on the market.

A more direct case of unfair competition comes to a head when creators wish to license their own creations for generative AI purposes or even release an AI model trained on their own creations.[26]

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)?

---

[22] DeepLearning.AI. (2023, June 7). LAION roars: The story of LAION, the dataset behind Stable Diffusion. *The Batch*. https://www.deeplearning.ai/the-batch/the-story-of-laion-the-dataset-behind-stable-diffusion/

[23] Beaumont, R. (2022, March 31). LAION-5B: A new era of open large-scale multi-modal datasets. *LAION*. https://laion.ai/blog/laion-5b/

[24] Such as LAION-5B, sourced from the Common Crawl, an internet scraper that does not filter out copyrighted works.

[25] In the sense of anticompetitive practices: https://www.ftc.gov/enforcement/anticompetitive-practices

[26] See an example of a vocal artist and musician who released a voice model: Herndon, H. (2021, July 13). Holly+ 👭🗣️💪. *Holly Herndon*. https://holly.mirror.xyz/54ds2IiOnvthjGFkokFCoaI4EabytH9xjAYy1irHy94

I believe opt-in should be the *de jure* approach regarding copyrighted works used as training materials.[27] It is fair to, respects, and rightfully empowers creators and copyright owners.

Opt-out creates (unpaid) labor and shifts the responsibility onto individual creators. Opting out imposes what I believe to be abnormal expectations on creators: requiring the labor and resources to individually opt-out on what may be a large body of their work, *multiplied* by the exploding quantity of AI models; and requiring the labor and resources to fight against corporate infringement in court, as a result of the forced assumption of responsibility on the individual. Were opt-out to be standardized, creators who are less tech-savvy (but nonetheless impacted) and *independent* creators (including marginalized voices) also cannot hope to compete with corporate copyright owners' abilities to opt-out on a massive scale.

The burden of opt-out also verges on the preposterous when considering that any copyright owner cannot even access the training datasets of proprietary models, cannot feasibly inspect the billions[28] of training points if granted access, and cannot feasibly input the thousands of personal photos and visual works into web services designed to perform such searches against AI training datasets.[29] A further, rather insidious issue surfaces from the latter point: the existence of intimately personal information—such as medical record photographs[30]—in training datasets from public scraping or automatically ingested as a result of "AI-enhanced" services that happened to have handled sensitive data.

A similar debate long haunts the fight for privacy amidst the routine, careless harvesting of personal data, with the same conclusion: we must have opt-in, *not*

---

[27] With the exception of research and educational purposes. See my response to question 9.1.

[28] Such as LAION-5B.

[29] Such as the website Have I Been Trained?: https://haveibeentrained.com/

[30] Edwards, B. (2022, September 21). Artist finds private medical record photos in popular AI training data set. *Ars Technica*. https://arstechnica.com/information-technology/2022/09/artist-finds-private-medical-record-photos-in-popular-ai-training-data-set/

opt-out.[31] As in the case of privacy rights, opt-out (as the *WIRED* article referenced mentions) gives technology companies plausible deniability—washing their hands of accountability over the rampant harm and infringement of the rights of copyright owners.

An article on California's "Delete Act" offers further insightful parallels: that opt-out (in the fight for privacy) is mired in dark patterns that obscure (privacy) concerns and discourage users from actively opting out, and that the process of opting out often involves transmitting (at times *additional*) verified personal information directly to the very same brokers who profit off of such data—"counterintuitive" indeed.[32] With regards to the 2018 California Consumer Privacy Act (CCPA), analysis has shown the ineffectiveness of opt-out forms: large percentages of data deletion requests were denied due to petitioners declining to complete identity verification.[33] A study conducted to ascertain efficacy of compliance with CCPA opt-out sheds further light on the problems encountered with opt-out.[34]

The same lack of efficacy will surely apply to opting out copyrighted works: already, opt-out forms such as OpenAI's form[35] commonly ask that artists bundle up the very images of their artwork with the hopes that they will be excluded from one set of training data. Of course, there are no guarantees this plea will be honored, no guarantees the volunteered images will not be directly sold to third-parties or transferred during an acquisition or used in-house for a

---

[31] Barrett, B. (2019, August 2). Hey, Apple! 'Opt out' is useless. Let people opt in. *WIRED*. https://www.wired.com/story/hey-apple-opt-out-is-useless/

[32] Sandle, T. (2023, September 1). What the California 'Delete' Act means for consumers. *Digital Journal*. https://www.digitaljournal.com/life/what-the-california-delete-act-means-for-consumers/article

[33] Luthi, S. (2021, August 5). 'Functionally useless': California privacy law's big reveal falls short. *POLITICO*. https://www.politico.com/states/california/story/2021/08/05/functionally-useless-california-privacy-laws-big-reveal-falls-short-1389429

[34] See this article for an even greater breakdown on the ineffectiveness of opt-out and actual user experiences: Waddell, K. (2020, October 1). California's new privacy rights are tough to use, Consumer Reports study finds. *Consumer Reports*. https://www.consumerreports.org/electronics-computers/privacy/californias-new-privacy-rights-are-tough-to-use-a1497188573/

[35] See https://share.hsforms.com/1_OuT5tfFSpic89PqN6r1CQ4sk30

profit, and no recourse should the aforementioned scenarios occur. The burden of proof would rest solely on the artists themselves.

Finally, opt-out does nothing to reverse the existing problematic AI models accessible on the market, whose myriad ill effects are listed extensively in my response to the first question. Machine "unlearning"—which can enact a retroactive removal of data from already trained models—remains a relatively open problem.[36]

On the other hand, opt-in fully respects the wishes and intellectual property of copyright owners and creators. It correctly shifts the burden *back* onto technology companies who possess the time and funding to ethically source their training datasets and compensate copyright owners who opt-in, as they should have from the start.

Specifically, I believe the right course of action is to take steps to ensure generative AI models (and, more broadly, AI/ML technology intended to turn a profit) with unauthorized copyrighted works in their training datasets are retracted—permitting only models that fully respect opt-in on the market and in the public.

Although opponents of opt-in are concerned this practice will impede progress in ML (perhaps only where such research has overt commercial application), I believe it imperative to course-correct; any delays are in actuality the fault of a lack of ethical oversight. This should have little impact on ML research (except, perhaps, where research was possible through funding by technology companies hoping for profits). Regardless, the rapidity of technological advancements our society experiences and the dizzying innovation we as a society are capable of lead me to doubt progress can be significantly impacted. Royalty payments (profit-sharing and the like) to individual creators can incentivize and properly compensate high-quality contributions from copyright owners and living creators.

---

[36] Pedregosa, F., & Triantafillou, E. (2023, June 29). Announcing the first Machine Unlearning Challenge. *Google Research Blog*. https://blog.research.google/2023/06/announcing-first-machine-unlearning.html

The requirements for opt-in must be carefully crafted so that they adequately serve copyright owners. The process must be free of "dark patterns"[37] and appropriately verify the authorization of the licenser. I would optimistically welcome opt-in that respectably and reasonably compensates creators.

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

With the limited exceptions of research and educational purposes[38], I believe that yes: consent of the copyright owner should be required for all purposes of training AI models or as inputs. However, models and systems developed in the course of research should not be automatically cleared for commercial or other purposes. In cases where text or data mining contributes to research, when non-research technology is developed, any and all such contributions should be purged of materials that would violate copyright laws. Commercial (and, generally, publicly accessible) systems should not reuse the training datasets, inputs, weights, or other information obtained from *direct* inference and contribution of copyrighted materials.

To do otherwise risks bolstering the "academic-to-commercial pipeline" and other "data laundering" tactics—to derive profit from the inevitable interval by which the law lags behind technology—currently at play.[39] Technology companies and their investors must be disabused of the notion that profit can be made from aggressive development ahead of regulations and ethical survey.

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

---

[37] See a breakdown on dark patterns here: https://www.eff.org/deeplinks/2019/02/designing-welcome-mats-invite-user-privacy-0

[38] Under limited circumstances, there may be artistic merit that falls under the scope of the *fair use* doctrine, permitting commentary or criticism that directly engages with the trained materials.

[39] See Andy Baio's article "AI Data Laundering: How Academic and Nonprofit Researchers Shield Tech Companies from Accountability" for a breakdown of data laundering and the academic-to-commercial pipeline: https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/

hire—should they have a right to object to an AI model being trained on their work? If so, how would such a system work?

I believe that human creators (irrespective of their ownership of the copyright) should indeed have the right to object with respect to their creations. However, I note that objection is still an opt-out system, and I continue to support opt-in as the *de jure* approach.

Despite whatever contracts were signed *prior to all parties' full knowledge of the implications of generative AI systems today*, I believe copyright owners who are *not* the original creator should *not* automatically inherit the right to opt-in or use the material in question for training AI models or as inputs. The exception to this, perhaps, are works released into the public domain (where nobody owns the copyright for that particular work, anyway).

Any work-for-hire contracts or transfer of copyright ownership should have a specific clause stipulating the right to use the material in connection with AI models.

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

In my response to question 9.1., I specified exceptions for the necessity of consent for training materials, viz. research and educational purposes. In particular, this allows for open-source research, which may, for instance, use publicly scraped data possibly containing unauthorized copyrighted works. (Whether such non-curated, large sets of data are always ideal for research is another matter.)[40]

However, I believe that outputs generated by AI models can still be infringing when such models contain unauthorized copyrighted works *and* the outputs are

---

[40] See §4 of "On the dangers of stochastic parrots: Can language models be too big? 🦜" by Bender, et al.

used for commercial purposes or publicly available in ways beyond for research/education (or, under limited circumstances, commentary or criticism directly engaging with the trained materials). For example, using an open-source model trained on copyrighted works is not itself infringing unless the outputs are then used in commercial settings or even directly released into the public domain. This measure intends to close off attempts to engage in data laundering.

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?

The law should require proper labeling and disclosure of the use of AI-generated materials. Services that offer AI generation must also make the necessity of disclosure prominently clear to their users.

The downsides of "black box" models—opaque computational systems that "use hundreds or even thousands of decision trees (known as 'random forests'), or billions of parameters (as deep learning models do), to inform their outputs"[41]—have long been documented[42] (Candelon, et al., 2023). Black box models, including many generative AI models—which operate across billions of training data—suffer from a lack of explainability.[43] How precisely are decisions reached? Which inputs are ultimately responsible and to what individual degree? What biases lie hidden in the data?

While the three previous questions remain to be answered by the researchers and engineers of such systems, I believe the law can aid "explainability" for the *public* with proper attribution and identification. Laws for proper disclosure are not unlike the spirit of the FTC's "truth-in-advertising" laws, designed to protect

---

[41] Candelon, F., Evgeniou, T., & Martens, D. (2023, May 12). AI can be both accurate and transparent. *Harvard Business Review*. https://hbr.org/2023/05/ai-can-be-both-accurate-and-transparent

[42] To clarify, black box systems in engineering have been discussed in academic literature for the past 50 years.

[43] *ibid*.

consumers from false or misleading information about commercial products[44]—particularly when the use of generative AI appears unexpectedly unannounced[45] or under inappropriate circumstances.[46]

ChatGPT—a text generator "chatbot" based off of a large language model (LLM)[47]—infamously elicits unwarranted trust from human users with its ability to generate authoritative, plausible prose, all while committing egregious factual and logical errors.[48] Due to the overly trusting tendencies observed of LLM users and the devastating (legal, financial, and perhaps even medical) consequences, I believe that laws of disclosure can make an actual positive impact.[49] Disclosure statements can nudge people to be more conscious of the content's potential for (seemingly authoritative yet abnormally extreme) errors, help people make more informed decisions regarding the content, and deter overreliance on generative AI (particularly where it is inappropriate).

AI-generated material—whether it includes human authorship or is entirely absent of human contribution—should feature the disclosure of such prominently alongside said material. The disclosure must clearly state the use of generative AI. Perhaps the additional requirement of disclosing the particular model used would be helpful, particularly when combined with legislation

---

[44] Federal Trade Commission. (n.d.). Truth in advertising. *Federal Trade Commission*. Retrieved October 29, 2023, from https://www.ftc.gov/news-events/topics/truth-advertising

[45] Oftentimes, the abrupt switch to using generative AI outputs leads to a sharp, noticeable drop in quality.

[46] See Sara Merken's article "New York lawyers sanctioned for using fake ChatGPT cases in legal brief" from https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/

[47] Language models (LMs) are described in the oft-cited paper "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜" as "haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot" (Bender, Gebru, et al., 2021).

[48] Goldman, S. (2022, December 5). The hidden danger of ChatGPT and generative AI | The AI Beat. *VentureBeat*. https://venturebeat.com/ai/the-hidden-danger-of-chatgpt-and-generative-ai-the-ai-beat/

[49] An example of an alarming consequence is documented in the aforementioned article, "New York lawyers sanctioned for using fake ChatGPT cases in legal brief."

requiring AI models to disclose training data sources—thus provenance can be tracked. Disclosures of this nature should appear in front of ("above" when considering scrolling text) the affected material, so that the general public can easily make the choice not to engage with it without needless searching. In the case of generated images, disclosure should additionally be embedded in the image's metadata.[50]

Finally, I would like to clarify that the inclusion of disclosures should not absolve the person or entity responsible for publishing or distributing the material from accountability.

[50] See §7 of the paper "AI Art and its Impact on Artists" by Jiang, et al.