US Copyright Office – Artificial Intelligence and Copyright

This piece was written up in response to the USCO's fielding of public voices and opinions on Artificial Intelligence (AI for short) and Copyright.

The bona fides:

Name: Thammakhun John Crowcroft
Portfolio: http://www.crowcroft.net/portfolio/
Job Role: High End Generalist in 2D/3D/VR/GameDev

I have already submitted my short comment previously to the 'AI Accountability Policy Request for comment' by the National Telecommunications And Information Administration. (Tracking# lis-ch9k-7u0b // https://www.regulations.gov/comment/NTIA-2023-0005-0998) Specifically concerning accountability and data/transparency practices around AI. This submission is a longer form comment regarding copyright and AI.

My comments here are submitted during the reply period to respond to the main arguments and points raised by AI companies and Trade groups:

In response to comments from:
Technet // Comment ID COLC-2023-0006-8767 // Tracking# lod-c20g-jgp4 //
https://www.regulations.gov/comment/COLC-2023-0006-8767

"Artificial intelligence may be the most transformative technological development since the creation of the Internet. Generative AI, in particular, has the potential to transform not only the creative industries, but other fields like software development, scientific research, healthcare, government administration, and education."

This is true, however, just because technology has potential, does not mean we discard any and all guardrails or ethics as to how we utilize and manage its use. If anything, the more potential technology has, the more scrutiny it requires in proportion.

"The technology, at its core, promises to enable all people to create new content in any medium or any language, regardless of their skill level or ability."

Enabling all people to "*create new content in any medium or any language, regardless of skill level or ability*", does not necessarily mean that any of that content will be particularly good, useful, or cogent. One can argue that enabling such ability. Far from being a banal and petty argument for special pleading for certain demographics or classes who 'deserve' to produce content, this is meant to address the inherent implications that this comment puts forward that such enabling necessarily IS a positive. Not even the underlying, deeper implications of who stands most to profit from the enabling of such abilities, much less at what costs and who incurs the negative externalities of such a transaction.

"By breaking down barriers to creative expression, research and communication, Generative AI will unlock a more robust exchange of ideas and a more robust information economy."

This is not necessarily the case. The advent of the internet has also, by the same token, lowered the barriers to entry to the propagation of information *and* misinformation. The high tide lifts all boats – both sturdy and rickety. Lowering the barriers to entry also enables entry to all sorts of actors – both good and bad. For the Tech group to only insist on the upsides shows that they are not inclined to even bother to think about the potential negative externalities of this paradigm.

"In this way, the promise of Generative AI aligns precisely with the goals of our intellectual property laws. This is a development copyright law should celebrate and encourage, not restrict. We urge policymakers to prioritize the preservation of core copyright law provisions that offer technology-neutral safeguards for legitimate rightsholders and innovators, which will enhance the competitiveness of American AI and creative industries on the global stage."

This is a non-sequitur. How does 'lowering the barriers' to entry ensure the promise that Generative AI aligns with the 'goals of our intellectual property laws'? If any thing the enforcement of barriers is precisely what

aligns with the stated goals of any regulatory restrictions ipso facto, *not* the lowering *or* raising. This is an anodyne talking point meant to disarm the critical thinking faculties of anyone reading the comment.

Training AI models does not inherently implicate rightsholders' copyright interests

"[C]opyright s not an inevitable, divine, or natural right that confers on authors the absolute ownership of their creations. It is designed rather to stimulate activity and progress in the arts for the intellectual enrichment of the public."1 Copyright, in other words, is a public policy tool to assure authors that if they invest time and resources in creating new works of authorship, the law will, for limited times, grant to authors certain exclusive rights to those works. This is not a "special reward" for authors, but a "utilitarian" incentive mechanism to "encourage the production of works" for the benefit of the public.2 In order to strike the appropriate balance between the rights of the copyright owner and those of the public, the Copyright Act not only grants exclusive rights to authors, but also imposes limitations and exceptions that limit the reach of those rights.

Just because copyright, "is not an inevitable, divine, or natural right that confers on authors the absolute ownership of their creations", it inversely does not necessarily follow that any there is free reign for people to abrogate said limited rights to begin with on a whim. TechNet is already starting to beg the question and presume from the outset that basic enforcement of copyright laws, and the stated goals and aims of copyright, already are 'absolute' and stifling the potential utility that the technology can possibly provide. Since TechNet themselves agree that the goal is a utilitarian incentive mechanism to encourage the production of works for the benefit of the public, their comments here can only be considered tautological.

"For example, copyright has never required a follow-on creator to pay or get permission from an author before studying and learning from the ideas, facts, concepts, styles, and themes in an author's work.3"

That's precisely because copyright does not protect copyright holders from others '*learning from the ideas, facts, concepts, styles, and themes in an author's work.*' – it protects the works and fixed expression that copyright holders invested the time and energy into creating and fixing. What TechNet here is starting to set up is the rhetorical baseline for anthropomorphizing of what is a mechanical process by using terms that are associated with human actions in order to linguistically contrive a false equivalence in order to circumvent copyright restrictions that would ordinarily apply otherwise, via use of 'studying and learning' of works that humans do, and applying that to a mechanical machine 'learning' process with the implication that since it's perfectly acceptable for humans to 'study and learn' from works, then it must follow that it is ok for machines to do so as well. This would be true if machines and humans were the same thing – the point that TechNet wishes to make in order for the rest of their argument to follow.

"In this way, copyright protects the ability of authors to market their works for profit, while also encouraging innovation, competition, and the free exchange of ideas. As the Supreme Court has recognized, "[t]he more artistic protection is favored, the more technological innovation may be discouraged; the administration of copyright law is an exercise in managing the tradeoff.

This statement is correct. It is true that copyright protections are not 'absolute' so they cannot stifle technological innovation, but inversely, the demands for technological innovation cannot stifle/abrogate copyrights.

"The goal of Generative AI is to help authors create new content—whether it consists of responses to user queries, longer pieces of text, visual works, music, or computer code. Encouraging the creation and dissemination of new expression, of course, is the very purpose of copyright law. No one could seriously contend that a technology whose sole function is to create new works is at cross-purposes with copyright law. The invention of the camera, for example, made it easier for people to create new visual works regardless of their skill with a pencil or paintbrush, opening the door to an entirely new and valuable medium of artistic expression."

This is preposterous. People CAN seriously contend that "*a technology whose sole function is to create new works is at cross-purposes with copyright law.*", IF that technology violates copyright/intellectual property law. This is not a question of what the technology is 'intended' to do, but HOW it accomplishes said function. In this comment, TechNet engages on confabulation by trying to muddy the waters between intent and methodology. Cameras do not fundamentally rely on copyright infringement of others works. The rudimentary pencil,

paintbrush, or photoshop does not. Those are tools that fundamentally, on a technical level, DO NOT require others fixed copyrighted works to produce new works.

"Similarly, Generative AI will make it easier for humans to create works in any medium, regardless of their skill level or ability,4 thereby fostering a more robust and democratic creative economy. This is a development that copyright law should celebrate, not restrict. To say that Generative AI is a threat to the copyright framework is to fundamentally misunderstand both the intent of generative AI technologies—to augment, not replace, creators—and copyright's central purpose: not to provide a "special private benefit," but to encourage "the free flow of ideas" and "creative activity."5"

TechNet's statement ITSELF, requires a fundamental misunderstanding of intent vs. functionality of generative AI technologies to lead into the abrogation of enforcement of basic copyright protections that the companies they advocate on behalf of stand to benefit the most from. This is the same incentive model whereby people who do not invest the time or effort into creating any fixed copyrighted works, stands to gain the most from infringement, and not respecting the basic copyright/property rights of others.

They are only 'similar' if you only look at what the function is and ignore the 'how' that functionality is accomplished, in which case they are not similar at all. Cameras or their operative use by any photographer does not require the fundamental use of others fixed copyrighted works, in either reference or on a technical level. Generative AI – as it currently exists that TechNet advocates on behalf of - cannot produce any outputs unless other copyrighted holders fixed copyrighted works are involved in the process at all. This is a fundamental requirement that technically does not exist, or applies, for the camera, pencil, or paintbrush. Therefore, the argument from TechNet this is a 'development' that copyright law should celebrate as opposed to restrict, at bare minimum, does not follow.

"Those who oppose AI on copyright grounds do not dispute these core principles. Instead, they raise two collateral attacks. First, some have complained that Generative AI will create works that compete with their own. These artists worry that the low marginal cost of creating AI-generated works—combined with the impressive capabilities of these models—threatens their ability to compete in the marketplace, and urge the Copyright Office or Congress to take some action to reduce that competition."

That's because the principles stated are not in dispute. The *admitted* fact by gAI companies that copyrighted works are being used, and thus, any form of competition is effectively copyright holders being forced to compete against their own works at low marginal cost of creating new works. That fundamental difference is what makes this a case of unlawful competition via intellectual property infringement vs. requests of special protectionism against lawful competition that did not involve intellectual property infringement.

"Copyright should not be used to restrict innovative technologies that may alter the competitive landscape by facilitating new modes of creativity. The use of copyright in this way would have condemned many valuable technological developments over the course of our history, including the invention of the camera (to which artists objected on the grounds that it would destroy art altogether7) and the advent of the modern recording industry (to which orchestra conductors objected on the grounds that it would destroy the demand for live performances.8) Those innovations have since enriched our creative economy many times over, both by opening the door to new media of expression and by generating massive new economic opportunities for creators.

Neither should the needs or demands of 'innovation' require the abrogation of intellectual property rights. The use of copyrights would not have condemned the technological advancement of the camera – or all the other examples of technologies that have enriched our lives - because the camera or those technologies literally do not require the fixed copyrighted works of others as inputs to function or produce anything. That's why those technologies exist today precisely BECAUSE they operate within the framework of copyright, in both law and spirit.

"Generative AI is no different.9 That the technology enables the creation of new works that might compete with the works of today's artists is not a reason to distort copyright into a tool of protectionism— particularly when doing so would have the collateral effect of impeding the use of Generative AI to solve some of humanity's "worst inequities"10 and most difficult scientific and medical problems.11 Current copyright law, in fact, does offer protection where there is true infringement, but to use copyright to throttle the creation of a technology for producing new works simply to protect existing copyright owners from competition would be to turn copyright policy on its head."

Except it fundamentally IS different, TechNet simply ignores the fact that generative AI technologies – as they exist – fundamentally requires access to others data – copyrighted or not – in order to function at all. If you remove access to others' fixed works, generative AI technologies could not function at all. No such equivalent technical stipulation exists for cameras, pencils, or paintbrushes to function.

Special pleading is fundamentally required to gain exemptions from basic copyright protections and restrictions that all other technologies and individuals must abide by whilst competing in the marketplace, in order for generative AI to function and exist. TechNet then engages in basic psychological projection of the very thing it is guilty of by accusing copyright holders of distorting *"copyright into a tool of protectionism"* when all copyright holders are asking for is the enforcement of basic protections of copyright. If anything, TechNet itself is attempting to overturn copyright policy on its head to cultivate special exemption/protectionism for itself to the detriment of copyright holders, by distorting the difference between intent vs. function of generative AI technto in order to obfuscate around the fundamental fact that generative AI fundamentally require the fixed work of copyright holders to function.

The test for this is very simple to demonstrate: If Generative AI technologies did not fundamentally require the fixed works of other copyright holders in order to produce anything, then, and ONLY then, would TechNet's argument follow. As it exists, it does not. Despite TechNet's insistence otherwise, a camera does not fundamentally require the fixed works of other copyright holders in order for a photographer to produce anything. Should they disagree, I'm sure myself and many other copyright holders would like to see TechNet's explanation as to how cameras and generative AI technologies work 'similarly' in this aspect.

"The second attack on Generative AI is formalistic. It hinges entirely on the fact that, unlike the process of human learning, today's Generative AI models must make intermediate copies of content in order to learn from them. But the creation of intermediate copies in furtherance of the creation of a new and useful technological tool is not the kind of copying that violates copyright law. In fact, courts and commentators agree that this sort of intermediate, non-expressive copying constitutes a fair use.13"

This is misrepresentation of what the courts have actually ruled, which would become obvious if you look at the wider court rulings on the cited cases and looked at the actual USE cases in question rather than just relying on specific cherry picked highlights from TechNet. The courts have not NOT ruled that *"this sort of intermediate, non-expressive copying constitutes a fair use."* When it comes to the act of intermediate copying, intermediate copying by ITSELF, does not 'constitute a fair use' as it still falls under the scope of copyright; Whether or not that specific use of intermediate copying is 'fair' depends on the use of what that intermediate copying is FOR.

"Courts have recognized this principle across multiple contexts involving innovative technologies, including internet search engines, which "make[] and analyze[] a copy of each Web page [on the Internet]" in order to produce useful search results;14 plagiarism-detection tools, which make unauthorized copies of term papers to spot cheating;15 video game console emulators, where copies are necessary to enable interoperability;16 and book search engines, where unauthorized copies of millions of books were made to facilitate new insights and keyword searching.17 Other countries around the world also recognize that copyright law must leave room for legitimate and useful text and data mining.1"

That's because, again, that is situation dependent on a case-by-case basis given what the use case in that case is for, and whether any act of infringement is considered 'fair' under the doctrine of Fair use.

With the indexing of web pages and search results, those uses of intermediate copying are considered fair as the purpose of indexing of web pages is to provide search results, and does not result in market displacement of the original copyright holder. To put it in simple terms: 'Search results' are not the same as 'The Actual website'.

With plagiarism detection tools, that's because again, the PURPOSE is for the detection of plagiarism and cheating, NOT copying essays in order to reproduce them for submission. To put it in simple terms: 'Cheating detection' is not the same as 'essay reproduction'

With book search engines, again, those unauthorized, intermittent copies were made to facilitate NEW insights and keyword searching. NOT replace the books as a whole and thus resulting in market displacement. To put it in simple terms: 'Keyword searching' does not replace the 'book'.

For TechNet's argument that Generative AI outputs are analogous to all the above, use of intermediate copying must result in a new work that has a transformative purpose or utility – something that has an entirely different use-case so as to not result in market displacement under the criterias for Fair Use. This is not the case as the intermediate copying process is used to create a product that is fundamentally of the same purpose/utility as the original copyrighted material is trained off of, at very low marginal cost, and explicitly for commercial mass scale purposes. Courts have already ruled that 'intermediate copying' of copyrighted materials for the purposes of creating mash-ups, do not constitute fair use. (https://www.copyright.gov/fair-use/summaries/drseuss-comicmix-9thcir2020.pdf) TechNet will be hard pressed to argue how a 'statistical interpolation process' of training data to produce generative outputs, that was acquired through the process of intermediate copying of copyrighted works, for mass scale commercial purposes – is somehow 'fair use' along the above cases. Considering the extent of market displacement that has already resulted from the advent of generative AI:

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

*"It is now possible for anyone to create hundreds of images in minutes, compile a children's book in an hour8, and a project for a successful Kickstarter campaign in a fraction of the time it takes for an actual artist9. Although many of these images do not have the full depth of expression of a human, **commercial image generators flood the market with acceptable imagery that can supplant the demand for artists in practice. This has already resulted in job losses for artists, with companies like Netflix Japan using image generators for animation, blaming "labor shortage" in the anime industry for not hiring artists [32].**"*

*"One of the more high profile cases of the labor impact can be seen in the title sequence of Marvel Studio's 2023 TV series Secret Invasion, which uses a montage of generated imagery [81]. While prior movies from the studio feature between 5 (The She-Hulk: Attorney at Law10) and 9 (Hawkeye11) artists and illustrators for their title sequences, Secret Invasion has only one "Sagans Carle" credited as "AI Technical Director"12. **This labor displacement is evident across creative industries. For instance, according to an article on Rest of World, a Chinese gaming industry recruiter has noticed a 70% drop in illustrator jobs, in part due to the widespread use of image generators [139]; another studio in China is reported to have laid off a third of its character design illustrators [139].**"*

Study Finds ChatGPT Drives Down Freelance Work, Pay - https://www.thewrap.com/study-finds-chatgpt-drives-down-freelance-work-pay/

"The release of ChatGPT led to a 2% drop in the number of writing jobs on one major freelance employment platform and a 5.2% plunge in monthly earnings, a new study found.

They decided to look at writing jobs after examining Google searches in the months after the release.

"'GPT Writing' is by far the most commonly search term compared to other domains, such as 'GPT translation' or 'GPT software development,'" they wrote. "We interpret these patterns as suggestive evidence, consistent with the notion that while multiple industries are affected by the introduction of ChatGPT, the short-term implications are much more pronounced in the writing-related tasks."

The decline in both the number of writing jobs and the compensation received via UpWork suggests "a causal relationship" with the release of ChatGPT, wrote researchers Xiang Hui and Oren Reshef of Washington University and Luofeng Zhou of NYU.

This shows that the resulting outputs from generative AI cannot be considered transformative due to having the same utility from the original copyrighted materials that generative AI makes 'intermediate copies' of, given evidence of already considerable, significant market displacement in such a short span of time in the

very same category of works/utility that copyright holders, as a class, now have to compete against works of the same utility/use, in the market place.

For TechNet's argument to be true, there should be no possible market displacement if the resulting 'new work' from generative AI was transformative enough for the use to then be considered 'fair' by virtue of add net benefit and utility to the public without causing economic harm to copyright holders. This is very clearly, not the case.

"Ensuring that copyright does not expand beyond "its lawful bounds" to become a veto over new technologies is one of the core purposes of our fair use doctrine.19 Indeed, as Judge Stephanos Bibas recently held, the intermediate copying of copyrighted content as a "step in the process of trying to develop a 'wholly new'" and non-infringing AI system is a "transformative" fair use, even if the resulting product "compet[es]" with the original.20 The well-established body of "intermediate copying" case law on which that holding relied21 has long been a bedrock part of our copyright framework that the Office itself has endorsed in its own Reports.22"

This is such a patently incorrect misrepresentation of what Judge Stephanos Bibas said that I believe that TechNet should be sanctioned for what amounts to essentially manufacturing a quote and opinion that literally does not exist. Their credibility should now be automatically suspect going forward given the stunning disingenuousness they have just displayed.

Here is what Judge Stephanos Bibas actually said in his Judicial Opinion of Thomson Reuters Enterprise Centre GmbH et al v. ROSS Intelligence Inc., No. 1:2020cv00613

317, at 11. And it contends that Ross merely translated the headnotes into numerical data and that translation is "paradigmatic derivative work[]." D.I. 317, at 10.

But Ross says its AI studied the headnotes and opinion quotes only to analyze language patterns, not to replicate Westlaw's expression. So the translation was only a minor step in a broader, transformative use. *See Sega*, 977 F.2d at 1514–15, 1518–19 (holding that, though programmers wrote down and translated Sega's object code, these acts were a minor step towards a transformative use). If Ross's characterization of its activities is accurate, it translated human language into something understand-able by a computer as a step in the process of trying to develop a "wholly new," albeit competing, product—a search tool that would produce highly relevant quotations from judicial opinions in response to natural language questions. This also means that Ross's final product would not contain or output infringing material. Under *Sega* and *Sony*, this is transformative intermediate copying.

So whether the intermediate copying caselaw tells us that Ross's use was trans-formative depends on the precise nature of Ross's actions. It was transformative in-termediate copying if Ross's AI only studied the language patterns in the headnotes to learn how to produce judicial opinion quotes. But if Thomson Reuters is right that Ross used the untransformed text of headnotes to get its AI to replicate and reproduce

19

"If Ross's characterization of its activities is accurate, it translated human language into something understand-able by a computer as a step in the process of trying to develop a "wholly new," albeit competing, product—a search tool that would produce highly relevant quotations from judicial opinions in response to natural language questions. This also means that Ross's final product would not contain or output infringing material. Under Sega and Sony, this is transformative intermediate copying." -

Judge Stephen Bibas did not 'hold' any such thing. He was setting up the hypothetical of "**If** Ross's characterization of its activities is accurate", not that it **IS**, much less that "*intermediate copying of copyrighted content as a "step in the process of trying to develop a 'wholly new'" and non-infringing AI system is a "transformative" fair use, even if the resulting product "compet[es]" with the original"* that TechNet claims Judge Stephen Bibas stated as a matter of legal precedent. TechNet's citation is manufactured and intentionally leaves out the full quote and caveat of "IF" that Judge Stephen Bibas said in his judicial opinion, in order to assert their characterization that intermediate copying is a "*step in the process of trying to develop a 'wholly new'" and non-infringing AI system*" is ""transformative" fair use" as though that is a matter of established legal precedent that has been determined via a trial, as opposed to him hypothetically characterizing The Defense's use case itself being transformative IF that were the case.

1) Which is exactly WHY the case is going to trial in the first place to determine the disputed facts of matter of the hypothetical IF that Judge Stephen Bibas was setting up. - https://www.reuters.com/legal/thomson-reuters-ai-copyright-dispute-must-go-trial-judge-says-2023-09-26/
2) That couldn't even be possible true as a matter of legal precedent, because if it was, the case wouldn't be going to trial IN THE FIRST PLACE.

TechNet's duplicity in the manufacturing of a quote from thin air to 'justify' their basis should, as far as the Copyright Office is concerned, render their credibility as 1) an honest actor and 2) competency in the matter, utterly questionable at best.

"AI companies have poured tens of billions of dollars of investment into Generative AI in reliance on that basic and fundamental point of law. That investment will yield huge advances in fields from physics to medicine to education, continuing economic growth, and tens or hundreds of millions of new jobs over the next two years.23 It will make countless people both more productive and more creative; it will certainly encourage, facilitate, and inspire the creation of more new, creative works than any technology since semiconductors and personal computers.24 The desire by a limited set of advocates to upend copyright's foundational principles for short-term revenue generation would undermine law's goal of advancing creativity and technology in the long-term."

It is not the responsibility of the US Copyright Office, or any other parties, to assume or pass on the liabilities of Venture Capital and Tech Firms that TechNet advocates on behalf of, who have poured billions into Generative AI, who are relying on a *"basic and fundamental point of law"* that they have manufactured out of thin air - by patching together quotes from a Judge in an AI copyright case concerning a hypothetical that he was setting up, that is still to go to trial, and shamelessly pretending that their citation is an established legal precedent and *"basic and fundamental point of law"* as opposed to an intentional misrepresentation they contrived up as a basis - to justify their otherwise baseless arguments, nor are they entitled to the US Copyright Office or anyone else, to pretend and agree with them that is the case. Nor is it the remit of the US Copyright Office to be swayed by Venture Capital's financial appeals that they've invested so much money into an area that they did not fully clarify was legal before doing so. And to thus 'allow' them to 'bend' policy towards in a way that is extremely lucrative to them. To do so would not only be falling straight into regulatory capture, but it would also set future precedent that instead of wasting time and money lobbying in Congress, firms can just simply wade into a new field, where the legalities have yet to be fully outlined, and use financially appeals to sway regulatory agencies into simply letting things slide. This would not only create an incredible moral hazard, but it would also render regulatory agencies obsolete.

If the Venture Capital and Tech firms that TechNet advocates on behalf of want to reap the tangible benefits & profits that the resulting technologies would bear fruit from, they should be expected to do so by complying with intellectual property laws instead of trying to openly flaunt them. No other existing class of firms or industries exist in the free market that do not have to pay for its factors of production. Why should Generative AI firms expect to have special protections from basic business requirements that no other firms have? Any 'potential upside' in the form of a boon of creation of many AI based jobs will be exceeded by the negative externalities incurred in the form of job losses due to the hollowing out of the intellectual property ecosystem that Generative AI firms will treat as the source of their raw material inputs that they have will have no incentive to compensate for if copyright protections were not enforced.

There is no basis to impose liability on AI developers for users' misdeeds

"For the reasons discussed above, copying material for training purposes does not automatically violate copyright law."

Again for reasons stated above, the 'intermediate copying' that TechNet claims "does not automatically violate the law" that they base the rest of their following arguments on, was based around a quote that they themselves manufactured, that does not factually represent any legal precedent nor what Judge Stephen Bibas actually said.

"The question of whether the outputs of generative AI systems may violate copyright law is a separate and distinct question, and one that is likely to turn on the specific facts of each case. That analysis, however, must begin with a prior question: if an output does infringe copyright, who should be responsible?"

Correct. Perhaps the only body of text in their entire comment that one can consider to be objective and impartial.

"Generative AI models, like many traditional software products, are general- purpose tools. Large language models, for example, can be used for a broad variety of tasks: to power analytical tools that analyze sentiment in a large set of online comments; to summarize or provide a critical review of a book; to translate speech from one language to another; to provide advice for a DIY project; to translate an idea for a short story into a piece of polished prose; or simply to engage in a conversation. The next generation of "multimodal" models—those that can understand and create outputs in any media, whether text, image, video, or sound— will have an even broader range of potential uses.25 AI models, in other words, will be "Everything Machines." The scope of their potential uses will soon be as broad and infinite as the uses of a programming language or an operating system."

"To impose liability on the creator of such a system for potentially infringing outputs makes no more sense than imposing liability on the creator of a word processor simply because someone used it to draft a work of fiction that infringes on a copyrighted book. In virtually all cases, it is the user of a Generative AI tool who supplies the prompt that dictates the resulting text or image. In the same way that the law might impose liability on an individual who uses Adobe Photoshop to create an unauthorized derivative of a copyrighted work (rather than imposing liability on Adobe for providing the tool), the law should impose liability on the individuals who use AI-powered image generators to achieve the same result."

This is incorrect. The legal basis for Vicarious Infringement is basic black letter law. For TechNet to pretend otherwise is either, again, a testament to their disingenuousness or incompetence.

The reason liabilities aren't imposed on the creator of a word processor is because there is no infringement done on part of the word processor (Or its creator) itself when a user of the word processor introduces copyrighted material into the program.

This is the same reason that liabilities aren't imposed on Adobe when a user uses Adobe Photoshop to produce an unauthorized derivative of a copyrighted work, is because the copyrighted material was introduced on part by the user, and NOT on the part of Adobe of Photoshop.

However, in virtually all use cases of Generative AI tool, the user who supplies the prompts that dictates the resulting infringing text or image, the infringement is necessarily possible at all in the first place because the copyrighted materials were introduced at all during the 'intermittent copying' stage. Which TechNet insists is a matter of settled legal precedent. (It isn't) But even assuming that was true as TechNet claims, then it would legally follow that even generating infringing outputs is fair use, as such, there would be no liability to impose in the first place; To assume to pass on any liabilities to begin with for Generative AI outputs as a use case necessitates that the steps there, the intermittent copying, cannot be considered fair use. TechNet essentially reveals there hand here that their intent is to hedge, obfuscate, and displace any responsibility for any costs or liabilities whilst seeking to maximize profits on behalf of the companies they advocate for.

"Imposing liability on the tool-creator would also be inconsistent with well-established copyright doctrine, which protects the providers of neutral tools from copyright liability as long as those tools are "capable of substantial non-infringing uses."26 Since the 1980s, that doctrine has protected "[i]nventors and entrepreneurs" from liability for the potential misdeeds of their uses, and has accordingly become one of the cornerstones of the pro-innovation environment to which we owe many of the technological developments that now power the American economy.27""

That is because, again, the creators of such tools have not done any intermittent copying of or introduced copyrighted works to even end up in any outputs in any tangible form at all, to be able to incur any liabilities. This is not the same as Generative AI. To NOT imposed liabilities on Generative AI tools that are infringing on copyrighted works, would absolutely be inconsistent with established precedents.

"This is not to say that developers of Generative AI tools should never be liable for infringing outputs—secondary liability doctrines may apply as they do to technology providers more generally. To be sure, because Generative AI is a novel technology unlike any other, and courts are best positioned in the first instance to develop new approaches for intermediating between "the respective values of supporting creative pursuits through copyright protection and promoting innovation in new [] technologies by limiting the incidence of liability for copyright infringement" in this new technological context.28"

Secondary liabilities cannot apply if there are no primary liabilities in the first place, which TechNet insists is/should be the case. I'm sorry, but *who wrote this*?

But the Office should not endorse stringent liability rules that force AI developers to incur liability for every user action. Such rules will impose additional financial barriers to participating in this field, thereby discouraging the "[i]nventor[]" in her "garage" or "dorm room" from entering the industry in the first place.29 They would also disincentivize developers from releasing foundational AI tools and models on an open-source basis, which, in turn, would make AI research far more costly for anyone other than highly-resourced laboratories and corporations. 30"

Again, this comment here reveals TechNet's hand that their intended goal is to hedge, obfuscate, and displace any responsibility for any costs or liabilities whilst seeking to maximize profits on behalf of the companies they advocate for. If there are confident that they are no liabilities to be enforced, whether they are enforced stringently or laxly would not be a concern by their own argument. Additionally, TechNet is arguing here that the 'barriers' that would impose 'economies of scale' – I.E. penalties to infringement of copyrighted materials which would otherwise amount to basic (licensing) cost of raw material data inputs - should be legally abrogated to encourage innovation whilst copyright holders/society incurs the cost of them doing so. That is not the remit of copyright holders, but of people who wish to innovate and reap the rewards from them. This would strike at the heart of what TechNet themselves state is the goal of copyright that is to ensure that there is public benefit in terms of utility, to the trade-off between copyright holders and innovation.

"Moreover, most of today's AI developers have undertaken careful measures to prevent or discourage misuse of models or the creation of potentially infringing outputs. TechNet encourages developers to continue and implement these responsible development practices. These measures have included contractual use- restrictions, technical measures like prompt- and output-filters, and warnings that models are not intended to be used to infringe copyright or other intellectual property rights. These are salutary and important steps that developers should continue to explore as the industry develops. But the possibility that a user might circumvent or ignore those steps and use models to for an unlawful purpose should not trigger liability for the developers who created the model."

This is patently false and absurd. If anything, the truth is the opposite. Only a minority of AI developers have undertaken such measures when the majority simply flaunt them. The reason for this is very simple: They have no incentive to do otherwise due to basic lack of enforcement mechanisms. Profit seeking entities do not have any incentive to engage in practices that are not profit-maximizing.

Opt-out requests are routinely not respected or followed through with, notwithstanding the onerous shifting of the burden of opting out to individual copyright holders.

OpenAI offers a way for creators to opt out of AI training data. It's so onerous that one artist called it 'enraging.' – https://www.businessinsider.com/openai-dalle-opt-out-process-artists-enraging-2023-9\

"OpenAI for the first time is letting artists remove their work from training data used for DALL-E 3, the latest version of its AI image generator. The opt-out process is so onerous that it almost seems like it was designed not to work.

OpenAI recently unveiled a new form that image owners and creators can use to request that owned or copyrighted images be removed from DALL-E training data.

AI models need high quality, and human generated training data to perform well. There's a race to accumulate all this information. But the original creators of this content have now realized that the value and intelligence embedded in their work is being ingested and processed for someone else's benefit. That's putting pressure on big tech companies to offer ways for creators to either actively decide to take part, or extricate their data from this grand AI experiment.

One by one

To have an opt-out request even be considered by OpenAI's new process, an artist, owner or rights holder has to submit an individual copy of each image they'd like removed from DALL-E's training dataset, along with a description.

...

OpenAI is full of very smart technologists. The company could have rolled out a process through which an artist or owner could make a single request that all of their work be removed from the training data. But the company did not do this. Why? Probably because it needs as much data as possible to build its AI models.

Too late

Even if OpenAI grants an artist or owner's opt-out request, it will only apply to "future" training data for DALL-E. The version 3 that was just released will have already made use of artistic work that a person requests be removed from its training data. Or, as OpenAI put it, its model will have "learned from their training data" and be able to "retain the concepts that they learned."

Translation: Here's the opt-out process, but it's too late because we've already sucked out most of the value from your work."

The fact that TechNet makes no mention of how the 1) onerousness of OpenAI's opt-out system, 2) how it isn't even retroactive, making the process utterly pointless, and 3) It externalizes the responsibility/cost of addressing the issue of copyright onto the copyright holder to enforce, with not even any guarantee or enforcement mechanism, or transparency laws/audits to ensure that OpenAI even abides by its own opt-out procedures, reveals, again, that their goal is the hallowing out of statutory and regulatory protections that are designed to safeguard copyrights, to maximize profits for the companies they advocate for at the expense of copyright holders and public utility.

That AI companies have "contractual use- restrictions, technical measures like prompt- and output-filters, and warnings that models are not intended to be used to infringe copyright or other intellectual property rights," in their terms of services again, would be irrelevant if they themselves were not infringing in the first place; Terms of service between the Firm and the User do not supersede, or insulate either, from statutory or regulatory requirements. No amount of prompting from an end user could or would amount to infringement technically if generative AI firms were not introducing copyrighted materials into the process in the first place.

Copyright does not extend to works that are entirely AI-generated, but line-drawing will be difficult

"The Office was correct in concluding that "copyright can protect only material that is the product of human creativity."31 As the Office is aware, however, the line between a work that is the "product of human creativity" and a work that is purely "AI-generated" is sometimes difficult to draw. 88 Fed. Reg. 16191. The good news is that these difficult edge cases are likely to be rare in general and commercially significant in even fewer cases. In these rare scenarios, we believe that existing case law is sufficient for the Office's registration analysis."

That's because, unlike the *"basic and fundamental point of law"* that TechNet believes is precedent re: Transformative use of gAI training, the fundamental human authorship requirement abundantly established legal precedent. For them to argue otherwise, or to persuade the USCO to consider otherwise, is a fool's errand and they know it.

"After all, most users of Generative AI will have no particular interest in the copyright status of most outputs, just as today nearly no one worries about the copyright status of their emails, instant messages, search results, and spreadsheets. And for those few creators who are interested in commercializing the outputs of Generative AI tools, there are numerous ways to combine their creative contributions with the contributions of an AI model in a manner that plainly yields a copyrightable final work. We encourage the Office to publicly state that human authorship utilizing Generative AI tools can be protectable under existing case law. For example, a human creator might use an AI model to edit, supplement, or improve a preliminary version of a work and then combine or arrange the AI-generated work with her own original creations. In these cases, the final work will support a copyright, giving the creator the ability to sell or license the resulting work and prevent infringing copying."

That there are lack of copyright protections given to generative AI outputs are precisely that the labour market for copyright holders has not completely fallen out from the bottom has that ensures that the labour market cannot be replaced wholesale by generative AI capabilities given the lack of exclusive monetization that copyright affords holders, for their efforts and investment. The

"This is not to discount the difficulties that may arise in the rare cases where distinguishing the AI-generated elements from the human-generated elements will rise to the level of commercial importance. In resolving them, time-tested principles of authorship that our courts developed far before the advent of Generative AI can be brought to bear. As the Supreme Court explained 140 years ago in a seminal case concerning copyright's treatment of photography:"

The US Copyright office has already given sufficient guidelines as to determine what is afforded copyright protections, and what isn't.

Statutory licensing schemes will not work in practice

"During the recent hearing before the House Judiciary Subcommittee on Courts, Intellectual Property, and the Internet, Register Perlmutter noted several concerns with proposals for a statutory licensing regime to govern the use of copyrighted content to train AI models—including, most prominently, that it was unclear how such a regime "can be made feasible given the volume of works that would be involved."33 This was entirely correct: while existing statutory licensing schemes might cover a universe of works in the tens of millions, a similar regime for works used for AI training purposes would need to administer royalty payments for many billions of works, representing many petabytes of data, most of which have no commercial value and were created without payment in mind. That would be orders of magnitude larger than any statutory or collective licensing framework ever deployed, covering a much more diverse set of works."

This is not a remit that the US Copyright Office, or copyright holders, need to be concerned about. This is exclusively the domain that Generative AI companies need to problem solve, and not externalize onto copyright holders/society just because the 'problem' is so onerous for them that it stands in the way of their profit seeking motive.

"That scale creates an intractable economic problem. Any licensing framework that provided any significant compensation to individual authors would impose a massive and insurmountable barrier to AI development, as it requires tens of billions of individual works—and, accordingly, tens of billions of individual royalty payments— to train an effective model. This would bar anyone other than the largest technology companies from participating in the AI industry, effectively foreclosing the next wave of American start-ups. Even the largest technology companies would have a strong incentive to take their AI development efforts (particularly for AI model training)— and the tens of billions of dollars of investment capital underlying those efforts—to other jurisdictions with more innovation-friendly legal frameworks."

Again, not the remit or concern for the US Copyright Office, copyright holders, or the rest of society. This talking point is perhaps the most ludicrous of all in TechNet's comment, and is rightfully deserving of opprobrium, mockery, and ridicule.

NO OTHER class of firm, in ANY industry, has the gall to complain that being 'required' to pay for its factors of production: From raw materials, to employee payroll, to overheads – would impose an 'intractable economic problem' for itself. If that statement is to be taken as true, then the only logical conclusion is that such a firm and business model by definition is not economically viable, and shouldn't exist. This admission from TechNet on behalf of Generative AI Firms is a direct admission that the Generative AI business model, by definition, is absurdly inefficient due to requiring its factors of production to be existentially subsidized – at the expense of not compensating copyright holders – and as such, could not possibly exist without it.

Subsidies, by technical definition, simply transfer wealth from one, more efficient sector of the economy into an underperforming, less efficient sector. With this admission, TechNet admits that Generative AI is not efficient, and thus undermines and contradicts any and all claims that any potential upside for the economy in terms of efficiency and job creation potential that Generative AI can bring. You cannot claim that Generative AI is economically efficient, much less viable or will bring an economic boom when it requires the hallowing out of the intellectual property ecosystem of said economic to even exist. The result is a net loss to the economy overall in terms of efficiency and health.

"On the other hand, any statutory licensing scheme that imposed a less crippling financial obligation on the next generation of AI developers would mean that the resulting payments to individual authors would be miniscule. To an individual author, there would be no meaningful difference between the current copyright framework—under which they have no right to collect rents for the use of their works to train AI models, see supra Part B—and a licensing framework under which they receive royalty checks of a few cents each month. Such a scheme, with its attendant inefficiencies, neither benefits creators nor promotes the progress of science and the useful arts."

TechNet begs the question and outright assumes from the onset that ANY *"less crippling financial obligation"* imposed by any statutory licensing scheme would result in miniscule royalties, when any determination on any royalties given would be the natural result and consequence of any free market negotiations that have been precluded from happening to begin with, given that copyright holders never had any say in the matter, and opt-out requests aren't even routinely enforced due to lack of incentive.

"Worse still, any statutory licensing scheme would be impossible to administer. As Register Perlmutter recently explained, such a licensing framework for AI training would need to cover "all kinds of works"— including not only commercially available books, musical works, sound recordings, and professional photographs, but also other in-copyright works like internet comments, blog posts, online reviews, social media posts, amateur photographs, etc. The Office has already noted how "time- consuming, difficult, or even impossible" it is to locate the copyright owners for literary works.34 However difficult it might be to locate the rightsholder for a book that has gone out print, that difficulty would pale in comparison to the challenge of identifying the rightful owners of the tens of billions of works used to train today's AI models."

Again, not the remit or concern for the US Copyright Office, copyright holders, or the rest of society. If TechNet and Generative AI firms want to explore the possibility and potential of Generative AI, and reap the profits and rewards, they must abide by the same rules that every other firm in every other industry must abide by.

"In this connection, it is also important to emphasize that, where training is concerned, AI models generally derive their value from learning patterns and relationships that emerge from the relationships between billions, or even trillions, of examples. In other words, in stark contrast to most licensing contexts, market values based on expressive uses are no guide for non-expressive training uses. The corpus of millions of online comments posted on a large user-generated content site, for example, is likely to be more valuable for training LLM models than the combined archive of a major newspaper. Of course, these UGC platforms do not own the copyrights in the content of their users, and thus would not be a proper recipient of any collective licensing proceeds. Accordingly, in fairness, any collective or compulsory licensing scheme would have to allocate most of the value to the individual internet users, rather than large commercial publishers."

This is factually incorrect. Whilst the statement that "AI models generally derive their value from learning patterns and relationships that emergy from the relationships between billions, and even trillions, of examples," that does not mean that all data is considered equal or that certain types of data aren't more, or less valuable. This is easy enough to establish given that Generative AI models cannot train on its own synthetic outputs without the models going 'MAD' and suffering from 'Model Autophagy Disorder' within the span of several iterative training generations.

Self-Consuming Generative Models Go MAD – https://arxiv.org/abs/2307.01850

Seismic advances in generative AI algorithms for imagery, text, and other data types has led to the temptation to use synthetic data to train next-generation models. Repeating this process creates an autophagous (self-consuming) loop whose properties are poorly understood. We conduct a thorough analytical and empirical analysis using state-of-the-art generative image models of three families of autophagous loops that differ in how fixed or fresh real training data is available through the generations of training and in whether the samples from previous generation models have been biased to trade off data quality versus diversity. Our primary conclusion across all scenarios is that without enough fresh real data in each generation of an autophagous loop, future generative models are doomed to have their quality (precision) or diversity (recall) progressively decrease. We term this condition Model Autophagy Disorder (MAD), making analogy to mad cow disease."

Researchers warn we could run out of data to train AI by 2026. What then? - https://theconversation.com/researchers-warn-we-could-run-out-of-data-to-train-ai-by-2026-what-then-216741

"Why high-quality data are important for AI

We need a lot of data to train powerful, accurate and high-quality AI algorithms. For instance, ChatGPT was trained on 570 gigabytes of text data, or about 300 billion words.

Similarly, the stable diffusion algorithm (which is behind many AI image-generating apps such as DALL-E, Lensa and Midjourney) was trained on the LIAON-5B dataset comprising of 5.8 billion image-text pairs. If an algorithm is trained on an insufficient amount of data, it will produce inaccurate or low-quality outputs.

The quality of the training data is also important. Low-quality data such as social media posts or blurry photographs are easy to source, but aren't sufficient to train high-performing AI models.

Text taken from social media platforms might be biased or prejudiced, or may include disinformation or illegal content which could be replicated by the model. For example, when Microsoft tried to train its AI bot using Twitter content, it learned to produce racist and misogynistic outputs.

This is why AI developers seek out high-quality content such as text from books, online articles, scientific papers, Wikipedia, and certain filtered web content. The Google Assistant was trained on 11,000 romance novels taken from self-publishing site Smashwords to make it more conversational."

Thus, to say that *"In other words, in stark contrast to most licensing contexts, market values based on expressive uses are no guide for non-expressive training uses."* is fundamentally incorrect. Even amongst organic, human data, there is a hierarchy in terms of quality of data that is more, or less viable, for Generative AI to train on.

"As a result, the amount of "unmatched" royalties under such a scheme would dwarf the amount of paid-out royalties by a significant margin. Most rightsholders would likely receive no remuneration whatsoever. Indeed, given the nature of AI training, which fundamentally is meant to extract facts and statistics from training data, developers can't determine how any particular work has been used, which makes any system of compensation that tries to value the contribution of each work fundamentally impossible. One of the most promising features of AI is that it finds relationships and patterns that we often cannot explain, much less find ourselves."

Again, that is not the remit of the US Copyright Office, or of copyright holders in general. Again, the statement that "***Most rightsholders would likely receive no remuneration whatsoever.***" Ignores the fact that 1) some data is, quite simply, more valuable that others; not all data is equal, and 2) no free market negotiation processes ever took place, they were simply precluded altogether. TechNet not only has no basis for asserting that "***Most rightsholders would likely receive no remuneration whatsoever.***" *to then say that "**developers can't determine how any particular work has been used, which makes any system of compensation that tries to value the contribution of each work fundamentally impossible,**"* again, is not the remit of the US Copyright Office, or of copyright holders: It is the responsibility of Generative AI companies to solve. If

"In summary, any legislation that would force AI developers to remunerate rightsholders for the use of their content to train AI models would ultimately fail to provide meaningful compensation to individual authors and, accordingly, fail to serve copyright's overall purpose of incentivizing the creation of new works. The only practical effect of a statutory licensing scheme would be to impose a substantial and burdensome tax on AI innovation—and, in turn, risk the transfer of American AI development overseas. This would in turn have severe detrimental impact on the long-term ability of creators and innovators alike to progress science and the useful arts as intended by law. U.S. Const. Art. 1, § 8, cl. 8. That kind of legislation has no rightful place in our copyright system."

Again, TechNet's summary here precludes the fact that:

1) Not all data is equal, some is inherently more valuable than others. So TechNet cannot extrapolate backwards from how Generative AI "***extract facts and statistics from training data***" to treat all data inputs as equal. That is logic is a non-sequitur and quite literally, counter-causal.

2) Any free market negotiations for licensing data never took place, so it is logically false and unreasonable for TechNet to claim that rightsholders would fail to derive any meaningful compensation from any licensing model. That is not their judgement to make or discern on behalf of copyright holders. TechNet can't have it ALL ways with:
i) How they claim Generative AI stands to bring immense value to the economy.
ii) That Generative AI can't exist without the corpus of data it requires to produce anything.
iii) But Rightsholders cannot be meaningfully compensated despite the value that their data will ultimately generate via AI.

This is a contradiction in basic logical syntax. If there is enough value generated, there is enough value to compensate rightsholders. The rest is pure laziness and unwillingness to properly adhere to the social contract that TechNet wants to derive benefit from, but not have to uphold.

The 'scare' that imposing this would be a burden that would risk AI innovation flight overseas, ignores that every other same jurisdiction has the same incentive from government and regulatory agencies to not allow the hollowing out of its intellectual property ecosystem just to feed an inefficient sector of the economy. Thus, this prisoner's dilemma that countries that don't adopt lenient regulatory frameworks for AI, will fall behind, is a false one.

Transparency obligations would unduly burden AI developers and handicap domestic AI development.

"In its Notice of Inquiry, the Office inquired about calls for copyright-adjacent legislation that would require AI developers to "disclose records regarding materials used to train their models" so that "copyright owners [could] determine whether their works have been used." 88 Fed. Reg. at 59947. That legislation would lead to a number of practical harms, as discussed below.

At a broader level, however, the fact that the works on which a model has been trained cannot be readily be determined by users of that model is proof that training works as designed—it is not meant to replicate the content of the works but to simply extract the unprotectable elements of those works."

This is preposterous and utterly irrelevant. Not only is whether or not the fact that works trained on **"cannot readily be determined"** is irrelevant to the enshrined right that the US Copyright act (https://uscode.house.gov/view.xhtml?hl=false&edition=prelim&req=granuleid%3AUSC-prelim-title17-section106) guarantees the *"(2) to prepare derivative works based upon the copyrighted work;"* being:

"Preparation of Derivative Works.-The exclusive right to prepare derivative works, specified separately in clause (2) of section 106, overlaps the exclusive right of reproduction to some extent. It is broader than that right, however, in the sense that reproduction requires fixation in copies or phonorecords, whereas the preparation of a derivative work, such as a ballet, pantomime, or improvised performance, may be an infringement even though nothing is ever fixed in tangible form.
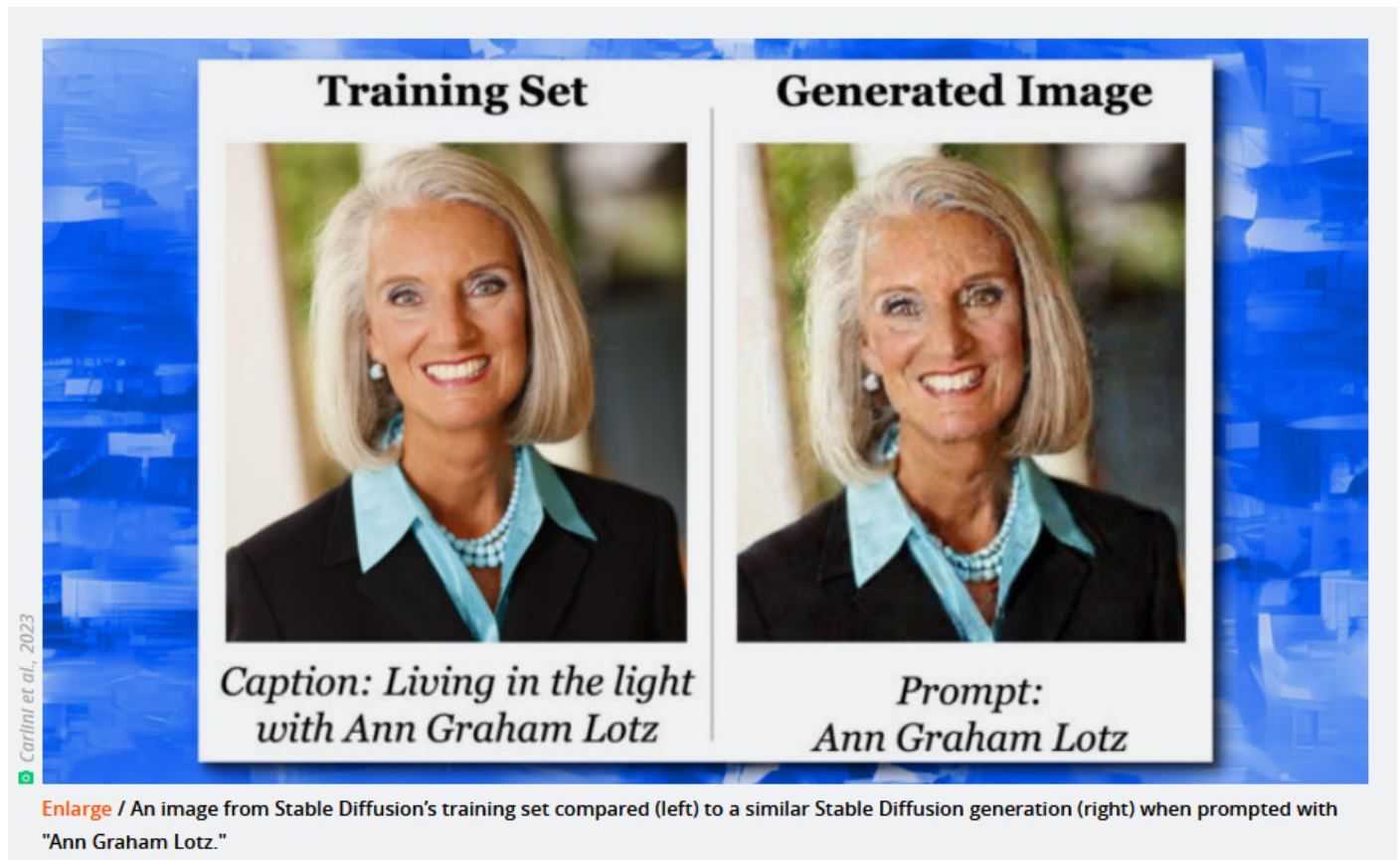
To be an infringement the "derivative work" must be "based upon the copyrighted work," and the definition in section 101 refers to "a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted." Thus, to constitute a violation of section 106(2), the infringing work must incorporate a portion of the copyrighted work in some form; for example, a detailed commentary on a work or a programmatic musical composition inspired by a novel would not normally constitute infringements under this clause."

One can VERY easily argue that, by the same token, that *"the fact that the works on which a model has been trained cannot be readily be determined by users of that model"* is proof that the training process was designed to obfuscate whether or not the fact that rightsholders works were used.

Not only that, regardless of whether Generative AI models are 'designed' to replicate contents of the works, the fact that it is technically possible at all ("Overfitting") demonstrates that on some level, copyright infringement that arises to the level of actual copying, is what happens underneath the hood when it comes to generative AI. So to claim that *"the fact that the works on which a model has been trained cannot be readily be determined by users of that model is proof that training works as designed—it is not meant to replicate the content of the works but to simply extract the unprotectable elements of those works."* is demonstrable nonsense.

Paper: Stable Diffusion "memorizes" some images, sparking privacy concerns -
https://arstechnica.com/information-technology/2023/02/researchers-extract-training-images-from-stable-diffusion-but-its-difficult/

Enlarge / An image from Stable Diffusion's training set compared (left) to a similar Stable Diffusion generation (right) when prompted with "Ann Graham Lotz."

"See supra Part B. We are aware of no other context in which a rightsholder has demanded remuneration for a use of a copyrighted work that is not only imperceptible, but impossible to detect.35 This is precisely why disclosure requirements have never been a part of copyright law: because the only uses of a copyrighted work that implicate rightsholders' interests are those uses that actually communicate the expressive elements of a copyrighted work, in some form or another, to an audience, thereby usurping the rightsholder's ability to do or authorize the same."

This is objectively false given even a basic black letter law of the US Copyright Act.

*"To be an infringement the "derivative work" must be **"based upon the copyrighted work,"** and the definition in section 101 refers to "a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, **or any other form in which a work may be recast, transformed, or adapted."** Thus, to constitute a violation of section 106(2), **the infringing work must incorporate a portion of the copyrighted work in some form**; for example, a detailed commentary on a work or a programmatic musical composition inspired by a novel would not normally constitute infringements under this clause."*

That's because the 'ability' of a rightsholder to demand remuneration based upon the perceptibility of expression ties into the judicial process of using substantial similarity to determine if infringement occurred within the contexts of civil litigation - NOT the black letter law requirement that enshrines the exclusive right of copyright holders to prepare derivatives – in order to plead to keep any requirements for data transparency, opaque, and to hamper any potential future civil litigation on copyright infringement where rightsholders would no longer have to go through the judicial process of substantial similarity to 'prove' that infringement occurred when data transparency would all but ensure and demonstrate that it did. In short, TechNet is conflating two different processes in order to plead ignorance as to be 'unaware' of rightsholders demanding remuneration for copyrighted works that are imperceptible or hard to detect, to circumvent basic data transparency requirements that would financially undermine the business model of the Venture Capital and Tech Firms that it advocates on behalf of.

"Should such a broad requirement be adopted, there is no justification for imposing it on only one industry. Movie studios, for example, frequently use existing copyrighted works to inspire, guide, and develop projects. Should a filmmaker disclose every pre-existing work that was passed around in email, shared via cloud file storage lockers, or scanned or photocopied during a film's development? Should musicians be required to disclose every poem read on the web, every mp3 sent in an email, or every Instagram cover art inspiration that played a role in developing a hit song? Should software companies be required to disclose every scrap of code that was consulted during their development of a new software system? In every one of these cases, the works that may have been copied are not readily apparent to potentially aggrieved copyright owners, unless the expressive elements of those works are detectible in the final product."

The premise here that TechNet relies on is the incredibly inept and lazy conflation between ideas, and protected copyright expressions of fixed works and also engaging in reductive anthropic fallacies in order to equivocate the way humans 'learn' ideas from fixed works vs. a mechanical processes that necessarily relies on works in its fixed form in order to actually produce anything. Are the copyrighted works themselves being used in these hypothetical scenarios that TechNet is postulation? Yes or No?

"Furthermore, legislation requiring broad disclosure of training data would discourage investment in critical AI technology and cripple American technology companies' ability to compete with their foreign counterparts. The creation of training datasets is a critical element of AI development.36

That is akin to saying that requiring broad disclosure of financial information regarding monetary transactions over $100,000 USD and basic KYC requirements will discourage investment in the financial services sector and America's ability to compete with their foreign counterparts when it comes to providing financial services of a more 'discrete' nature to clients that operate with a very relaxed attitude to the rule of law and/or illicit activities. Yes. That's precisely the point. TechNet here is, again, begging the question. Or that disclosing to the US patent office as to how the underlying technology actually works (In order to receive protection from the US Patent Office and US Government) will "*cripple American technology companies' ability to compete with their foreign counterparts*".

"Companies who build AI models devote substantial resources to selecting datasets that will yield effective, unbiased models."

This is literally not how the act of indiscriminate web-scraping works. This statement is objectively false.

"Which data is licensed, how it is obtained, and from whom, constitutes competitively valuable information. Still more resources are devoted to filtering data to (for example) remove duplicates and scrub potentially sensitive information, and to pre-processing data into a format suitable for model training. As a result, the "materials used to train [AI] models" are not merely a grab-bag of copyrighted works, but rather a highly curated set of data."

Indeed, it is decent of TechNet to at the very least acknowledge that there is economic value to the data they use whilst making the case that they should not have to compensate rightsholders for it whilst at the same time causing economic displacement. The fact that "**more resources are devoted to filtering the data",** or that the *"materials used to train [AI] models" are not merely a grab-bag of copyrighted works, but rather a highly curated set of data."* '- even if true - is irrelevant and not the remit or concern of the copyright office.