Questions from Senator Tillis
for Karla Ortiz (08/08/23)
Witness for the Senate Committee on the Judiciary Subcommittee on Intellectual
Property Hearing "Artificial Intelligence and Intellectual Property — Part II:
Copyright"

1. *Given generative AI is developing all over the world and countries are responding to it in different ways, are there policies or regulations being adopted elsewhere that you recommend that the U.S. consider or avoid?*

The U.S. has the world's most vibrant economy for creators, including writers, artists, designers, and software programmers. I am among the millions of Americans who've built a career in that creative industry. If generative AI companies plan to participate in that economy, they need to comply with the same laws that the rest of us do. So far, they have not.

First and foremost, I believe that dataset transparency in generative AI (currently being considered in the EU) is a vital policy for keeping AI companies accountable. If AI companies are required to disclose their datasets, they will have an obvious incentive to make sure their data is acquired legally and ethically.

Further, I do not believe that the U.S. should enact any statutory exemption for AI training, akin to the Text and Data Mining exemption in the EU. Experience has already shown that generative AI companies will find ways around it. For example, certain AI companies are already abusing the exemption through a practice known as "data laundering", where for-profit companies exploit a training exemption supposedly restricted to research. The fair-use doctrine under U.S. copyright law suffices for any generative AI company devoted to research.

I also wish to list again some of the policies that I think would benefit the U.S. greatly:

   a. Congress should update the Copyright Act to reaffirm that copyright requires human authorship.

   b. Congress should empower existing agencies to regulate the use of data to train generative AI models. This can take the form of requiring disclosure of training data, limiting the types of data that can be used to train AI models, closing "research to commercial" loopholes and ensuring compliance with these regulations.

   c. Congress should pass laws expressly authorizing those who have had their data used to train AI models without their consent the

right to vindicate those rights in federal court and to seek statutory damages. This can take the form of an amendment of the Copyright Act to authorize an express civil cause of action for those who have their data used to train AI models without permission. This can also take the form of passing a law authorizing a federal civil right of publicity cause of action.

d. Immediate taxing of all companies that replace jobs with any form of generative AI.

e. Immediate protection of citizens' data that powers generative AI, including likenesses, voice, biometric data, private data, copyrighted works and so on.

f. A copyright registration should be deemed sufficient for copyright-infringement litigation as long as the registration is gained before trial.

2. *A recent survey on how consumers view AI found that most consumers – nearly 80% – believe the use of AI should be explicitly disclosed. Do you agree? Why or why not?*

With the caveat that I have no plans to use AI for my own work, I think an AI-disclosure rule may still be insufficient. The problem with an AI-disclosure rule is that it is possibly too blunt a tool. It may make sense today, when works can be readily classified into either human- or AI-generated. Over time, however, this will become difficult, as more work could be a combination of human and AI contribution. What amount of AI contribution needs to be disclosed? How can that even be measured or verified? And what accountability will there be for those who violate the disclosure rules to try to pass off their AI work as human-made? I believe the better policy is to make generative AI systems legally accountable at the source.

3. *What are the benefits and disadvantages of requiring an AI company to keep records of everything that is ingested and to make those records publicly available?*

I strongly believe that dataset transparency is a critical policy for ensuring legal and ethical accountability of generative AI companies and the systems they create.

First, it creates a strong incentive for these companies to adopt legal and ethical practices for data collection and licensing.

Second, it's better for the generative AI industry at large, because customers of these systems will want to know that the outputs of generative AI systems are safe, legal, and ethical to use in other settings.

Third, it allows creators and artists to discover whether their data has been scraped for use in generative AI models. This third point is particularly important because in the case of large language models, the scraped data might be private or personal, and thus potentially affects every citizen with even the smallest digital footprint.

The U.S. has long imposed analogous labeling requirements for food and medicine for similar reasons. Disclosure requirements enhance safety and create confidence.

> a. *Under what circumstances, if any, should an AI company NOT be required to make its records of everything that is ingested by the AI publicly available?*

AI companies should always be required to make these records available.

> b. *Under what circumstances, if any, should an AI company be required to make its records of everything that is ingested by the AI publicly available?*

These records should always be made available. In a practical sense, what every generative AI system is offering is a certain kind of access to the underlying training data. All the "intelligence" in the system is derived from that training data. So it is conceptually impossible to frame the outputs of such a system as being separate from the inputs.

4. *Do you think that generative AI prompts provided by users are copyrightable? And if so, under what circumstances could they be copyrightable?*

That is primarily a question for the U.S. Copyright Office. If a generative AI prompt is, say, a sequence of words, then the copyrightability of that sequence should be judged by the same standards as other textual works. I would note that the current AI image generators, such as Stable Diffusion and Midjourney, are introducing features to make prompting easier, in response to complaints from users that even the current text-prompting systems are too difficult and complicated. It seems inevitable that these systems will move in the direction of supporting simpler prompts, not more expressive ones.

> a. *Do you think that whether the prompt used is copyrightable or not should impact the copyrightability of the resulting AI output generated as a result of the provided prompt?*

This is also primarily a question for the U.S. Copyright Office. In my view, copyrightability of a prompt and its output are two separate issues. As a human artist, when I make an oil painting, all the creative decisions originate in my brain and my hand. Whereas in a generative AI system, the person providing the prompt is supplying only a tiny fraction of what we might call creative agency. Mostly, the algorithmic decisions are being made by the generative AI system itself. So we have to consider the prompt separately from the output. Otherwise, we'll be led to absurd results. For instance, if the prompt were a copyrightable text like "provide an exact copy of an existing oil painting by Karla Ortiz / if you please" and the system produced such an image, then the copyrightability of the prompt obviously should not make the resulting infringing image copyrightable. This would also set a deeply unfair standard of essentially allowing AI users and AI companies to "launder" copyrighted works: anyone could type a prompt describing a previously copyrighted work and then claim copyright on both the prompt and the AI output. This act would single handedly destroy the need for anyone to respect copyrights.

5.   *What does the impact of generative AI have on the creative industry? Specifically, what are your thoughts regarding the concern that the proliferation of generation AI will take over jobs?*

As someone who works as a concept artist in the entertainment industry, it's apparent that the loss of jobs is not merely a "concern." A key reason generative AI systems exist is to replace human labor. IBM, for instance, has announced that it already plans to replace approximately 7,800 human jobs with AI. Emad Mostaque, CEO of Stability AI, has openly predicted that AI will have a "bigger economic impact than the pandemic." OpenAI has published a paper concluding that language models will have a significant impact on jobs, especially higher-income positions. Two major entertainment-industry creative unions, the WGA and SAG-AFTRA, are currently on strike in part because of disagreements with studios about the appropriate role for generative AI.

In my particular industry, every day brings more accounts of students and professionals seeing their projects reduced or losing opportunities altogether. I personally have been a part of three productions where generative AI has materially lessened my duties. As a freelancer who gets paid hourly, it has already impacted my bottom line. I expect that freelance concept artists and artists in the commercial-entertainment industries have and will continue to experience declining job assignments. I expect that there will also be downward pressure on pricing, and many of us will just have to find some other way to support ourselves and our families as the work dries up.

6.   *If a generative AI system is found to infringe a copyrighted work, who should be liable for the infringement – the AI company, the user providing the prompts to the AI tool, or both?*

Both, depending on where in the process the infringement is happening. Infringement can occur a) during the training of the AI system, b) as part of the distribution and operation of the AI system, c) when the AI system is generating outputs, and d) possibly other points.

7.   *In your opinion – currently or in the foreseeable future – can AI generated material ever replace the quality of human created work?*

No, it cannot. And what's more, AI companies agree. Human-generated content is indispensable as training data because it contains the widest variety of expression. Training an AI system on its own output leads to a condition known as "model collapse" where the system starts to "forget" things and thereby becomes less valuable. Because of this entanglement, it's not a stretch to predict that if the AI companies bankrupt human creators, they'll just end up bankrupting themselves. Thus, ensuring the health of the human creative industry should not be seen as a cost imposed on the AI industry, but as an investment in its long-term health and prosperity.

8.   *A balance needs to be struck in terms of how to encourage innovation, how to be responsible, and how to ensure that there is clarity for all using this technology. How do you propose we do this in the copyright space in a way that allows the U.S. to stay competitive and remain the global leader?*

During the last 50 years, U.S. copyright law has moved almost uniformly in one direction: toward greater protection for authors and creators. Why? Because as our national economy has become more reliant on industries rooted in copyright, affirming these rights has become vital. It's no coincidence that the last 50 years have also been the most prosperous in U.S. history. But now, the nascent AI industry is nudging Congress to reverse direction and erode protections for copyright. This makes no sense. If the generative AI opportunity is as big as these companies claim, then they should have no objection to complying with the established copyright regime that has been so successful for so long. Copyright law has **never** been a hobble to U.S. global competitiveness—on the contrary, it has been crucial to creating jobs and wealth that simply don't exist in other countries. Let's build on what works, rather than tearing it down.

9.   *In the copyright context, what differentiates the technology of generative AI from other machine-aided creativity, such as photography, video cameras, electronic music, and the like, all of which allow the public to develop and advance knowledge?*

There is a critical, fundamental difference between generative AI and previous technologies: generative AI systems are trained on vast quantities of existing copyrighted work, and representations of these works are retained within the AI model. The value of a generative AI system lies specifically in its ability to recreate expressions found in that

work. This is not true of, say, a video camera, or an image-editing program like Photoshop, which by default contain no representations of copyrighted works.

10. *What steps can and should the creative community take today to ensure that their work is more easily attributed to them, regardless of whether their work is used for training an AI model? For example, indicating authorship and contact information via the metadata of the author's digital content.*

I don't think new burdens should be imposed on creators to defend their works from AI training. Burdens like these would be contrary to the long-term policy of U.S. copyright law, which over time has eliminated most procedural formalities for achieving copyright protection. Under current U.S. copyright law, those who want to use copyrighted works have the burden of finding the author and asking permission. That is fair. Some creators may not want to make that easy! That should be their right. I, for instance, would prefer that my work is never used for AI training, and copyright law grants me that right.

11. *Are existing laws and regulations sufficient to deal with the issues relating to transparency and record keeping by AI companies?*

Probably not. As late as 2022, AI companies tended to see their work as being rooted in both commerce and research. They routinely disclosed details of their training methodology, including the training dataset, to help advance the AI industry as a whole. Starting in 2023, these companies have put up new walls, becoming increasingly cagey and opaque about these facts. Creators and artists are not asking AI companies to be subject to different principles of fairness than, say, food or pharmaceutical companies in terms of disclosing the ingredients that go into their products. But the techniques for implementing these principles may necessarily be different, given some of the unique features of generative AI technology. I do believe, however, that regulations and laws need to urgently clarify that these companies must be transparent, they must be thorough with their recordkeeping, and they must publicly share records of their training data.

12. *Have you reviewed the U.S. Copyright Office's Registration Guidance for "Works Containing Material Generated by Artificial Intelligence" and, if so, what are your views on the guidance?*

Yes, I have reviewed this document. In general, it seems the U.S. Copyright Office is applying the human-authorship requirement to AI-generated works in a way that is conceptually and legally consistent with historical practices and U.S. law. (Bearing in mind, as the Copyright Office notes, that "AI-generated works implicate other copyright issues not addressed in this statement," such as the possibility of copyright infringement during the process of training, deploying, and operating these systems.). I believe the Copyright Office has made a correct decision that will protect human creators and the jobs they rely upon.

a. *Do you think that the Copyright Office got it right? Are there aspects of the guidance that could stand to be clarified or revised?*

The U.S. Copyright Office's registration guidance says that "[w]hen an AI technology determines the expressive elements of its output, the generated material is not the product of human authorship." I thoroughly agree with this conclusion, and hope this basic idea, that authorship belongs to humans, becomes the foundation of any further action or policy.

13. *Both the U.S. Patent and Trademark Office and the U.S. Copyright Office have engaged in extensive outreach regarding AI. Have you participated in this outreach and, if so, how did you find it? What more can and should these offices do?*

Yes, I have participated in discussions with the U.S. Copyright Office as a speaker for their listening sessions in the Visual Arts category. I personally found the U.S. Copyright Office to be wonderful and professional. I do have a few suggestions for the U.S. Copyright Office that would be helpful to the discussions and also helpful to creative professionals like myself.

For starters, I believe that every governmental conversation and/or exploration concerning generative AI should represent all viewpoints fairly, and not be tilted in favor of industries poised to materially benefit from generative AI. For example, at my recent panel appearance at the Copyright Office, three artists (including myself) provided one perspective, while two generative AI industry representatives and three lawyers all spoke in favor of generative AI. While I think the listening sessions still went well, I believe better care in who gets to speak in these events should be taken. For instance, inviting more technical and academic experts—who have not been employed or funded by AI companies—to have a voice in these discussions and help determine policy. Further, those whose data is the subject of training (for example, artists, coders, filmmakers) and those who may be the subject of training (for example, educators, healthcare workers, etc.) should be represented in these discussions and have a role in determining regulatory policy.

I would also love to see the U.S. Copyright Office streamline copyright registration by making batch registration possible for artists. Currently artists like myself do not enjoy the same privileges as photographers, who are able to register mass quantities of their work all at once. Making this small shift would immediately benefit my artist community, and allow us to gain affordable access to the additional legal protections that copyright registrations provide.

14. *Can you describe some of the challenges that you face as a working artist and how you expect generative AI technology to impact those challenges?*

Generative AI would fundamentally change the lives of working artists. Consider generally the process and steps it takes to deliver finalized personal visuals to a client: A client describes what they wish to see. I, as the artist, send a variety of sketches to showcase potential visual avenues. The client shares their thoughts on the sketches, and we make adjustments until we feel it's ready to serve as the foundation for the final version. Once time is spent polishing, the final version is sent to the client. Generative AI replaces the artist in every one of those steps. It essentially eradicates the need for an artist. It does that work faster, cheaper, and delivers more quantity than any human artist could ever do. And all of this is based on exploiting the artists' work in the first place as training data.

It's important to note that artists have different industries they make a living in, each with their own challenges. I speak as an artist who works specifically in the commercial entertainment industries. Working as an artist in any industry, let alone the entertainment industry, is already an immensely challenging endeavor. While the entertainment industries are the only ones where an artist can find stable and consistent employment, these industries are notoriously difficult to get into and with good reason. For starters, the number of jobs available are already very small. Furthermore, the high level of technical ability required to gain these jobs, the time needed to make the necessary connections to enter these industries, and the high demands of these jobs, all contribute to the industry's difficult reputation. Due to the specialized nature of some of these jobs, wages do tend to be high (though they have gone down throughout the years). Because of the nature of our jobs, as described above, I see generative AI to be in a unique position where it can take the few jobs that are already available.

I expect generative AI to massively impact our industries. I could see a future where there is only one visual director editing the errors out of generative AI in a production, where normally there would have been a team of dozens of artists.

If left to proliferate, generative AI will make an already difficult living impossible. I genuinely do not see how my industry will survive.

15. *Do you feel that the value of your work has been negatively impacted by the advancement of these new generative AI systems? If so, how?*

I am proud of the work I do. I have trained my entire life to be able to have the technical ability and understanding to create art the way I do. I have enjoyed a successful career with many accolades. I never once doubted my future as an artist—until now.

Generative AI has affected the value of every single artist I know, in foreseen and unforeseen ways. The skills we worked so hard to gain have now been automated. In a cruel twist of fate, that automation was only possible because our own works were taken —without consent, credit or compensation—to train generative AI systems. With generative AI, at the click of a button you can make digital replicas of artists, you can also generate countless images at quantities never before seen, and you can do so via a low monthly subscription. No singular human artist can ever hope to compete.

Because of this exploitative novelty, artists, myself included, have experienced a sudden shift in our industry. We have a harder time negotiating livable wages when there's cheap software out there that can do our jobs. We have quietly seen an almost immediate devaluation of the skills we offer. Although no official studies have been made, reports of generative AI's impact amongst peers at all levels in our industry continue to surface. For instance, I've heard accounts of legendary veterans in my industry asked to paint over AI generated imagery, essentially asking some of the best artists of our time to no longer paint the evocative imagery countless have seen and simply fix a visual error here and there for less pay. I've also heard accounts of students losing out on internships because the position was now utilizing AIs, entry-level jobs being canceled, regular clients suddenly disappearing, usually busy times now being dry.

While the value of an artist is immeasurable, while we are all proud and confident in what we do, market forces simply do not agree. Especially with such a formidable cheaper and faster alternative to the costs of employing an artist. Again, if left to proliferate as is, being an artist may be something only a select few will ever be able to do.

16. *While style is not protected by copyright, do you believe that allowing for style to be copied and for users of AI systems to include artists' names in prompts results in a greater chance that the output will be substantially similar to an ingested work and potentially infringing?*

First—we should be clear that "substantial similarity" is not an element of copyright infringement. (See *Range Rd. Music, Inc. v. East Coast Foods, Inc.*, 668 F.3d 1148, 1154 (9th Cir. 2012).) Rather, it is a special evidentiary rule that helps courts determine if copyright infringement occurred in cases where a work *was not* directly and entirely copied. But when a work is directly and entirely copied—for instance, during the training phase of a generative AI system, where millions or billions of works are directly and entirely copied—then a finding of copyright infringement is automatic. Because a generative AI system can only "remix" expressive elements from the training data, there is a plausible argument that every output of that system infringes on the copyrights of the owners of the training data. Though some reject this idea as being impossibly broad, it is consistent with a quantitative account of how these systems work.

Second—style is indeed not copyrightable. But we should be careful to distinguish the two meanings of the word "style". In one sense, the word "style" denotes general moods, art movements and approaches—such as "realistic", "surreal", "fantasy art" and so on. As an artist, I have learned from the style of other artists, and I hope others learn from mine. But the other meaning of the word "style" refers to how I, as an artist and individual, uniquely approach and create *my* work. The only reason the name "Karla Ortiz" within a prompt means anything to a generative AI system is because it has been trained in part on *my* work—without consent, credit, or compensation. Furthermore, certain AI companies have promoted the ability of their systems to generate works in the style of certain artists, such as myself. Even if my style isn't copyrightable (though my work certainly is), I ought to have a right to prevent my name from being commercialized in this manner—which has also happened without consent, credit, or compensation. Thus, as a complement to copyright law, I hope Congress will consider a federal right of publicity law that will help artists and creators ensure that their names and reputations are not misused by AI companies.