

Questions from Senator Tillis
for Dana Rao

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 United States has led the world through technological transformations in the past and we have all learned that it is important to be proactively responsible about the impact these technologies can have on us. Pairing innovation with responsible innovation will ensure that AI ultimately becomes a transformative and true benefit for society.

At Adobe, we believe a comprehensive analytical framework for encouraging responsible AI development should consider:

- the impact of unfair AI bias on high-risk uses of the technology;
- the importance of ensuring access to data to spur the growth of AI;
- the impact of AI on creators and IP;
- the economic impact of AI on productivity and job creation; and
- mitigations against the spread of deepfakes.

We believe AI regulation should appropriately categorize AI systems as high risk and low risk and focus review on those high-risk systems. Examples of high-risk areas include AI systems that make decisions about health, employment, housing or finance. Any regulation or review should be conducted on a use case basis (versus categorizing AI as general purpose and applying the same standards regardless of use case). This will ensure that we are focusing on mitigating risk for those high-risk use cases, without unnecessarily slowing down innovation, so companies can continue to develop cutting-edge technologies and maintain U.S. leadership in the global technology industry.

Additionally, transparency in digital content is critical, especially in the age of AI. Congress should support provenance solutions and standards for providing consumers with more transparency about the content they are consuming – such as the Content Credentials technology and its underlying [C2PA](#) standard. Congress should also require any product or platform that receives content with provenance metadata attached to preserve and display that metadata (and not strip it away).

And as discussed in our written testimony, AI is trained on data, and training on larger datasets helps ensure the AI is more accurate and less biased. This is why we have seen other jurisdictions such as Japan, China, the United Kingdom, and the European Union recently taking steps to liberalize copyright laws and creating text and data mining exceptions specifically for AI/ML training. However, one of the important implications of AI's need for data is the impact on copyright and creators' rights.

We believe there is a way both to protect creators' rights and also to ensure the access to data that AI requires. The U.S. must balance the two to maintain a long-term leadership position in this space.

To protect creators, we encourage government and industry to work together to:

- provide artists a "Do Not Train" tag to allow them to opt out of training
- provide artists a way to secure copyright and attribution in a world with AI-assisted works
- establish a new federal right to allow artists to protect themselves against people misusing AI to impersonate their style
- and help advance new economic opportunities for creators enabled uniquely by AI.

Pairing these creator protections with support for access to data for AI research will ensure AI innovation continues to develop in the right way, for everyone.

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?

Yes. At Adobe, we believe transparency in digital content is critical, especially in the age of AI. This is the mission behind the Adobe-led [Content Authenticity Initiative](#) (CAI). The CAI is a group of more than 1500 members from across industries working to help enable creator attribution and fight misinformation through open standards for provenance technology and supporting tools like Content Credentials. Content Credentials allow creators to attach important information to a piece of content like their name, date, and what tools were used to create it. That information travels with the content wherever it goes. For generative AI, Content Credentials can show you whether a piece of content was human-created, AI-edited or AI-generated. In Adobe's own AI image-generating model, [Firefly](#), we automatically attach Content Credentials that indicate that content was AI-generated. This level of transparency in digital content will help consumers make more informed decisions about whether to trust the content they see online and help create a more trustworthy digital space.

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?

- 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?

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?

At Adobe, all of our AI products and features are developed in accordance with our [AI Ethics principles](#) of accountability, responsibility, and transparency. As part of our commitment to transparency, we share the types of datasets used to train the generative AI models we build and release in a commercial product. For example, the first version of Adobe Firefly was trained on licensed images from our own Adobe Stock photography collection, openly licensed content, and public domain images where copyright has expired.

We understand that practically speaking it can become very difficult to understand what datasets are used in commercially licensed models or open-source models, or datasets that are an amalgamation of other datasets. In addition, AI models may be trained to dynamically update with user interaction data, in which case the complete source of training data will be difficult to quantify or characterize. Accordingly, we suggest the requirements of disclosure may be best limited to references to named datasets (if any), or accurate characterization of the types of data on which the model is trained (e.g., user data, data licensed under a particular license agreement, copyrighted data, data in the public domain, etc.).

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

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?

Under US Copyright law, you cannot copyright an idea. What you can copyright is the expression of an idea. For example, you cannot copyright the word “book” – that is an idea. But you can copyright a painting of a book because that’s an expression of the idea. When you apply this principle to image-generating AI, it means that a prompt may not be copyrightable because the prompt represents the idea, and the output is based on the AI’s interpretation of that prompt. When you type in “cat driving a 1950s car through the desert”, the AI decides whether it’s a Siamese cat or a Tabby cat; it decides whether to make the car a convertible or a pickup truck; it chooses the color of the sky, the number of cactuses in the desert and so on. The AI is expressing the idea, not the human. Copyright law is designed to protect the rights of human creators – therefore an AI output (the AI’s expression of the idea) may not be copyrightable.

However, most creators will not use the raw, unmodified AI output as their final creation. Many creators are ideating and brainstorming in a generative AI tool and then adding their own style and expression to an image. Some artists are using AI just to quickly change the sky color of their artwork, rather than laboring through a tedious pixel-by-pixel manual task. Take our cat in a car in the desert example. Maybe the artist starts with the AI-generated output but then uses other

non-AI tools to change the color of the car, add some mountains in the desert, add another cat in the passenger seat, and so on. Or, maybe they already created the image of the cat in the desert using non-AI tools and they used AI to simply swap out one of the cactuses for a desert tree. In these instances, the human creativity and expression they are adding to their work should be copyrightable, but they will need to prove what was AI-generated and what was human-created. Fortunately, there are tools (such as any that implement Content Credentials) that will allow a creator to distinguish the portions of a work created by AI versus the portions of a work expressed by a human.

Currently, we do not believe that the text prompts typically used are of sufficient specificity to approximate the expression of an idea – they are more akin to an idea itself – and therefore may not be copyrightable. However, as the space evolves, it is possible that prompts may take different forms, and could become of specific precision and detail such that parts of the output could be considered the creator’s expression of an idea.

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?

We believe AI holds vast potential to unlock new opportunities for creators. First, we believe that Generative AI will increase productivity by automating tedious or repetitive tasks, allowing creators to do in seconds what used to take hours and increasing the creative output of every creative professional. Second, by making creating easy and fun, billions of latent creators can now become creative, because the tools are easy to use. As more people create, we believe the demand for professional content will only continue to grow leading to more opportunities for creative professionals. Third, we believe that creativity will be one of the more difficult aspects of human thinking for AI to imitate. In the creative world, we believe consumers of art prefer their art to have a soul and a story along with it in order to create the real connection and meaning that viewers, readers, and listeners want from it. This should mean that the creative professions are in a better place to be safeguarded from AI replacement. All of these points should encourage the United States to invest in creative skills training, to create a workforce equipped with the higher-level abstraction skills needed to differentiate against AI tools.

We also believe the AI business model itself can unlock new economic opportunities for creative professionals, and Adobe is exploring ways to help creators monetize their work in the age of AI. One approach would be to enable creators to license their style directly to consumers. In this approach, people could subscribe to a particular artist, who has worked with an AI model to have it replicate their style. The consumer could then create their own work in that artist’s style, but for a fee. This would be an entirely new revenue stream for artists, and we are currently testing an AI model with creators to do just this.

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?

Questions about copyright infringement should center around the output of a generative AI system and whether a particular output infringes on the copyright of someone whose image was used to train an AI system that generated that output and would depend on very specific facts of the case.

We believe generative AI companies have a responsibility to take steps to mitigate potential infringement, as do users of the tool.

For example, we believe that in cases where someone is intentionally using AI to try to impersonate an artist's likeness or style, the person misusing the AI tool in that way should be held accountable. This type of misuse can cause economic harm to artists. In terms of whether a tool itself should be held accountable, this should typically be decided on a contractual basis based on the user and the provider. In Adobe's case, we are indemnifying enterprise users for the works generated directly by Adobe Firefly. Other AI providers may choose different approaches, and customers should be free to choose the model they like.

To help enable artists to enforce their rights against misuse, we believe Congress should establish a federal anti-impersonation right (FAIR) that would give artists a right to enforce against someone intentionally and commercially impersonating their work through AI tools, without having to rely solely on copyright. This new law should include statutory damages to alleviate the burden on artists to prove actual damages, directly addressing the unfairness of an artist's work being used in a manner that could cause a creator direct economic harm.

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

As noted above, in the creative world, we believe consumers of art will want a soul and a story along with the art they are experiencing, to create the real connection and meaning that viewers, readers, and listeners want from their art and the artist that created it. In addition, we believe the evolution of style takes human creativity.

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?

Placing thoughtful safeguards around AI development and use will help us harness the full potential of AI to benefit society. We believe there are important steps that industry and

government can take to protect creators' rights in the age of AI while ensuring the U.S. continues to lead the world in AI development.

We believe government and industry should leverage the Content Authenticity Initiative's Content Credentials technology. Content Credentials are built on an open standard and have several important capabilities when it comes to protecting creators:

- **Do Not Train.** First, we believe that government and industry should support enabling creators to attach a "Do Not Train" credential in the metadata of their work. This gives them the option to keep their data out of AI training datasets. The Adobe Content Credential enables Do Not Train tags as part of its open standard.
- **Enabling creators to obtain IP for AI-assisted works.** Second, creators using AI tools want to ensure they can obtain copyright protection over their work in this new era of AI assisted digital creation. We believe that AI output alone will not receive copyright protection, but we believe the combination of human expression with AI expression will and should. With Content Credentials, creators can capture the tools they used throughout their creative process and distinguish their own non-AI efforts from the AI expressions in their work. This will give them the proof they need to obtain a copyright.
- **Enabling creator attribution.** Third, it's very easy to reproduce digital content in different forms and in different outlets, and the original author attribution can often get lost or stripped away. Content Credentials allow creators to cryptographically associate their identity with their work so they can get credit for it, wherever it goes.

We also believe it is important for the law to protect artists against the economic harm that could be caused by AI-generated works. We propose that Congress establish a new federal anti-impersonation right (FAIR) that would protect artists from someone using AI to impersonate their work or style. This new law would allow artists to protect their livelihood from people misusing this new technology, without having to rely solely on copyright.

We also know that to create good and unbiased AI, you need a lot of data. Like the human brain, AI learns from the information you give it. If AI has more facts to learn from and a wider breadth of experiences to build its experience of the world around it on, it will be more likely to generate accurate responses that avoid perpetuating harmful biases.

It is important to pair creator protections, like we proposed above, with support for access to data for AI research, whether it is through fair use or other means, so we can ensure AI innovation continues to develop in the right way, for everyone.

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?**

Like other technological advances in the creative space, such as advances in photography, video cameras, electronic music, etc., consumers are excited about generative AI as a new technology that has the potential to unlock vast new opportunities for creators. Generative AI differs from traditional machine-aided creativity because it allows direct creation of a work typically through a text interface. In the past, such as in Photoshop's own AI tools, the AI features worked on existing content. However, we believe in essence both types of AI are still machine-aided creativity. The creative professional will use the output of generative AI as the first step in their creative process, and will refine the work to match their vision. The Generative AI approach does, however, significantly advance the professional's ability to reach their end state much faster than before. In addition, we expect Generative AI will be used by creative professionals in all kinds of new and innovative ways, pushing forward the fields of innovation and the arts in unforeseeable ways.

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.

In the digital world, it is very easy to reproduce digital content in different forms and in different outlets, and it is often hard to maintain the source attribution for the original work. The Adobe-led Content Authenticity Initiative is a global coalition working to fight misinformation and enable creator attribution through open standards for provenance technology and supporting tools like Content Credentials. With Content Credentials, creators can cryptographically associate their identity with their work (for example, an image or a piece of digital art), and then if the work is reused or repurposed, their identity will travel with their work across all platforms that support Content Credentials.

To support this approach, Congress should require that all platforms that receive such attributions maintain them, so the attributions are not stripped away and artists can receive credit for their work.

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

As governments look to establish standards and safeguards to help guide the responsible development of AI, they should begin by leveraging existing frameworks, such as the NIST AI Risk Management Framework. Adobe collaborated with other industry leaders to help develop this framework. We believe AI governance should be global in nature and build upon industry best practices.

This framework outlines the characteristics of trustworthy AI, which includes transparency "about an AI system and its outputs." Adobe strongly agrees with the need for transparency

when it comes to AI. As stated in our AI Ethics Principles, and as discussed above, transparency means we are open about how we use AI and the types of datasets we use to train our AI. We are committed to working together with our community to design and implement AI that respects our customers.

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?

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?

Adobe has reviewed the U.S. Copyright Office’s Registration Guidance for “Works Containing Material Generated by Artificial Intelligence” and we are aligned with the approach the Copyright Office has taken. Adobe Senior Director, IP & Advertising Law, J. Scott. Evans, recently participated in a [U.S. Copyright Office AI listening session on visual arts](#), where he shared Adobe’s perspective on copyright in the age of AI and subsequently shared our perspective in this [blog post](#) for the Copyright Alliance.

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. Adobe Senior Director, IP & Advertising Law, J. Scott. Evans, recently participated in a [U.S. Copyright Office AI listening session on visual arts](#), where he shared Adobe’s perspective on copyright in the age of AI. We are encouraged to see these offices seeking feedback and taking a collaborative approach and Adobe is happy to continue to share our perspective and provide feedback in these ongoing discussions.

14. Scraping the Internet for data – text, images, audio, video, etc. – for use in training AI models has all the current focus. However, once this has been done the focus may shift to sources of data that are not as readily accessible, such as private user data.

Do you foresee companies using cloud-based file storage systems – such as Microsoft OneDrive, Google Drive, Dropbox etc. – as a potential source of data to be scraped? What are your thoughts on this?

We do not feel qualified to speculate about what other companies may or may not be doing.

15. Can you explain the thinking behind Adobe’s decision to adopt an opt-in approach to the use of copyrighted works as training data?

a. Are you concerned that this approach will put Adobe at a competitive disadvantage with those who do not see the need to obtain consent?

In designing the first model of our own family of generative image AI models, Adobe Firefly, we chose a path that supports creators and customers by training our model only on licensed images from our own Adobe Stock photography collection. This provides us with a dataset that is designed to minimize exposure from legal concerns.

However, it's important to remember AI is only as good as the data on which it is trained, which is why the question of data access is important to companies and organizations building foundation models. Like the human brain, AI learns from the experiences or information you give it. And like the human brain, the more information you give it, the better it will perform. An AI system trained on a small dataset is more at risk of producing wrong or unsatisfactory results, or reproducing harmful biases that exist within the dataset.

If you had never been taught what a car is, it would be hard to accurately depict one or answer a question about what it is or what it does. To produce accurate results, AI needs a large dataset representing the universe of possible answers to learn from. Additionally, a narrow dataset can lead to unfair bias. If you have only ever been taught that lawyers are men, you are likely to conjure up an image in your head of a man when someone is talking about a lawyer, even though more than half of the graduates of law schools are women. AI works the same way. Training on a larger dataset can help ensure you capture a broader set of perspectives in the data itself, so that when you type in "lawyer," you will see a result set that reflects the society in which you live.

We believe that designing the model this way mitigates the risk of infringing on someone's intellectual property, which allows enterprises and individual creators to create confidently with it. However, because we trained on a narrower dataset rather than scraping the web, our approach required extra engineering effort, testing, and mitigation practices to ensure accuracy and reduce bias in Firefly outputs. We encourage the government to support access to data to ensure that AI innovation can flourish both accurately and responsibly especially in areas where a large dataset of licensed data is not readily available.

16. Some of the most popular generative AI tools were built by indiscriminately scraping material – including copyright protected material – from the Internet. Can you explain Adobe's approach to building its AI tools and how it's different than other companies?

Adobe recognized the various unanswered legal questions around access to data in designing our own family of generative AI image models, [Adobe Firefly](#), which we launched in March 2023. We chose a path that supports creators and customers by training on a dataset that is designed to be commercially safe, where users have protections against potential legal liability that could be caused by other types of training, as the law on what is permitted is not currently decided.

We trained our first Firefly model on licensed images from our own Adobe Stock photography collection, openly licensed content, and public domain images where copyright has expired. To help ensure copyrighted or branded materials are not created as part of Firefly's output, we have a content moderation team that performs extra filtering on the images before they become part of the Firefly dataset. By designing the model with licensed content, the resulting model is more insulated against legal exposure from potential copyright claims.

- 17. One concern about generative AI that has been raised by creators is that unauthorized copies of their works are being made during the process of collecting data and training a respective model.**

Could you please explain how copies and how many copies of such data are made and when within the lifecycle of creating and executing an AI system – from start to end?

Adobe used licensed data, and not any unauthorized data, to train the first model of Firefly, our text-to-image generative AI tool.

- 18. Some have suggested different licensing structures for compensating copyright owners for the use of their works in AI training. What licensing structures have you seen or used that have worked to the mutual benefit of both AI companies and copyright owners?**

Adobe trained the first model of Firefly, our text-to-image generative AI tool, only on licensed images from our own Adobe Stock photography collection. However, it is still in early days of addressing the question of creator compensation and we have not yet seen a license structure announced that appears to address this. However, as we discussed in our testimony, we believe that we do need to enable access to data while giving creators a way to address the economic dispossession caused by a person misusing AI to impersonate their work. We believe this model will most directly address the harm caused by AI impersonation.