



**KARLA
ORTIZ**

Concept Art
Illustration
Fine Art

Written Testimony of Karla Ortiz
US. Senate Judiciary Subcommittee on Intellectual Property
“AI and Copyright”
July 7th, 2023

Chairman Coons, Ranking Member Tillis, and esteemed members of the Judiciary Committee, thank you for inviting me to testify today about AI and Copyright. “AI” stands for “artificial intelligence.” But that’s a misleading term, because, in fact, these so-called artificial intelligence systems depend entirely on vast quantities of copyrighted work made by human creators like me. These AI companies use our work as training data and raw materials for their AI models without consent, credit, or compensation. Today, I would like to share with you why these practices represent a dire threat to many American jobs, including mine. I will also explain why making AI fair and ethical for everyone is essential to building a sustainable relationship between AI and copyright.

My name is Karla Ortiz, a concept artist and illustrator based in San Francisco. You may not know my name, but you and your loved ones have almost certainly seen my work. My work has helped shape and create the worlds of blockbuster films including Marvel projects like *Guardians of the Galaxy Vol. 3*, *Loki*, *The Eternals*, *Black Panther*, *Avengers :Infinity War*, and *Doctor Strange*. I have also submitted art and designs for video games such as *Final Fantasy 16*, and trading card games like *Magic: the Gathering*. I am passionate about my work as an artist. Thanks to my hard work, my passion for the arts, and the support of my community, my work has been internationally recognized, and my fine art has been presented in countless galleries both domestically and internationally. There are hundreds, maybe thousands, of concept artists and illustrators like me who work tirelessly to make the films, TV, video games, and other media that bring entertainment to millions of Americans every day.

I am no longer certain of my future as an artist—a new technology has emerged that represents an existential threat to our careers: generative artificial intelligence (“Generative AI”). Generative AI is unlike any tool that has come before, as it is a technology that uniquely consumes and exploits the innovation of others. As I will explain in more detail, Generative AI relies on data it ingests in order to train models and generate its output, and oftentimes, that training data is the work of creative people like myself, taken without our consent, without any credit, and without any compensation. Based on this training data, Generative AI is capable of generating facsimiles of creative people’s work, including my own. As a result, AI has grown rapidly—in 2022, research estimated the value of the global AI market at approximately \$100 billion USD, and is expected to grow to nearly \$2 trillion USD by 2030. No other tool has the potential to adversely affect the rights of artists and creators. No other tool solely relies on the works of others to generate imagery. Not Photoshop, not 3D, not the camera, nothing comes close to this

technology. I should know, I've worked with almost all those tools. And even though artists and creatives are among the first American workers who are confronting the threat posed by AI on the job market, we are certainly not the last. Generative AI threatens the livelihoods of an untold number of Americans.

I am not fundamentally opposed to Generative AI. But AI needs to be fair, and ethical for everybody—and not only for the companies that make AI products. AI needs to be fair to the customers who use these products, and also for creative people like me who make the raw material that these AI materials depend upon. These systems depend entirely on the work of humans, especially creatives such as visual artists, writers and musicians.

The most important thing I can tell you is something the AI companies would prefer to overlook: "artificial intelligence" is really just human intelligence, repackaged and made available in a high-tech package. These models are useless alone, and need to be trained on an immense amount of data, including the work of artists of all creative fields. This is because generative AI is not truly "intelligent"—a Generative AI model cannot create anything it has not already seen in its training data. And therein lies the problem—the companies that make AI products have decided to neither ask for permission nor ask for forgiveness. Rather than ask artists and creatives their consent to use their work to train their Generative AI models, AI companies have instead opted to ingest massive amounts of copyrighted and licensed data to use as training data without permission, without crediting the creators of the underlying work, and without any compensation.

As a result of their wholesale ingestion of ill-gotten data, AI companies have reaped untold billions in funding and profit. Unsurprisingly, the AI companies have assured everyone that what they are doing is fair, ethical and legal. But the artists who made the works that their AI's rely on have never been asked for their consent, have not received any credit, let alone any compensation. In any other sphere, these practices would offend basic principles of fundamental fairness. And while the AI companies claim that what they are doing is fair use, but this claim has not yet been litigated. But the courts are beginning to weigh in. In May, a federal judge sustained important parts of a complaint filed by coders challenging the use of their code as training data for Generative AI models without regard to the requirements of the open-source licenses that code was subject to.¹ I myself have brought a lawsuit against Midjourney, Stability AI and DeviantArt for the use of my art as training material for their AI products.² While the courts have offered an important counterbalance against the interests of these big technology companies by providing an avenue to attempt to safeguard or vindicate important rights and to bring attention to this important issue, lawsuits may not be enough.

As I mentioned before, even though artists may be the first people to be affected by AI, the practices employed to develop these Generative AI models can be adapted to any kind of human expertise. Today, it might be me, who designed a loved superhero; tomorrow, it could be nurses, teachers, accountants, doctors, architects or any number of professions.

In our system, it is the role of Congress to pass laws, and to empower agencies to develop regulations to help our society deal with changes in time and technology—for example, ensuring AI companies have

¹ Order Granting in Part and Denying in Part Motions to Dismiss, *Doe 1 v. GitHub, Inc.*, Case No. 22-cv-06823-JST (N.D. Cal. May 11, 2023), ECF No. 95.

² *Andersen v. Stability AI Ltd.*, Case No. 23-cv-00201 (N.D. Cal.).

artists opt-in to having their works included as training data, laws that offer artists more robust methods of vindicating their rights or ensuring all AI companies to be completely transparent in their training data and contain no copyrighted works, biometric data or private data is utilized without right holders explicit permission. t. It is my hope that my testimony will help you to promulgate common sense laws that will make AI fair and ethical for all.

I. My First Encounters with Generative AI & Lessons Learned

My first encounter with generative artificial intelligence art models occurred a little over a year ago in April 2022. I stumbled upon a website called Weird and Wonderful AI Art. This website featured countless images alongside the names of artists, some of whom were dear friends of mine. I realized all the images were synthetic media, or generative AI images. These images claimed to be “studies,” attempting to make these synthetic images look just like the work of my peers. My initial reaction was one of interest. I was curious, so I reached out to my friends whose names were featured on that website. None of them were aware of the website, however. In fact, not a single artist had been asked to be a part of this “study”. When we tried to reach out to the folks who were running that website, to ask them to please remove the artists who did not want to participate, **we got ghosted**. Little did I know at the time that this would only be the first of many similar experiences I’ve had with generative AI.

In August and September of 2022, larger Generative AI models like Midjourney, Stable Diffusion and DALL-E are now more mainstream. In part because of my prior experience with generative AI, I did some research—I was horrified by what I found. I found that almost the entire body of my work, the work of almost every artist I know, and the work of hundreds of thousands of other artists, was taken without our consent, credit or compensation to train these for-profit technologies. I found out that once trained our work could not be forgotten. And to add insult to injury that these for-profit companies were not only permitting users to use our full names to generate imagery, but encouraging it. Some artists’ have had their names used as prompts a staggering number of times. For example, names of artists like the Polish artist Greg Rutkowski, had his name used as a prompt between Midjourney, Stability AI and the porn generator Unstable Diffusion, about 400,000 times as of December 2022. (And these are on the lower side of estimates).

I could not believe that an entire industry that so uniquely relied on ill-gotten data had suddenly emerged and was profiting without any regard to those it took data from; artists, creators and other rights holders . I connected with leaders in the AI/machine learning space to confirm what I had learned. Each and every AI/ML expert I talked to, from the founder of the Montreal AI Ethics Institute, to esteemed professors at the University of Chicago,³ to acclaimed researchers at the Distributed AI Research Institute were shocked at how exploitation fueled these synthetic media generators. Through speaking with these experts, I learned that due to the immense amount of data needed to power these models, it was very likely that every **single model** in the market right now contains huge amounts of copyrighted or otherwise licensed data. For example, the more popular models such as Stable Diffusion and Midjourney utilize a dataset created by LAION, a non-profit itself funded by Stability AI, which contains 5.8 billion image and text pairs taken indiscriminately from the web. LAION includes almost all of my fine art work, almost all of the work of my peers, the copyrighted works of my client without any regard to the rights of the artists

³ I myself worked alongside the researchers at the University of Chicago to develop “Glaze,” a system designed to protect human artists against generative ai’s style mimicry. <https://glaze.cs.uchicago.edu/>.

themselves Reporting indicates that these models may also be trained on other private data, such as medical records, nonconsensual porn, and violent images.⁴

Of course, Generative AI models such as Stable Diffusion and Midjourney are hardly the entire Generative AI industry. Other industry leaders in the Artificial Intelligence/Machine Learning (“AI/ML”) media sector likewise obscure the sources of the data that fuels their models. But despite the secrecy, it has been revealed that AI companies of all stripes rely on copyrighted materials taken without permission in order to train their models. For example, researchers at the University of California, Berkeley have found that Open AI’s Chat GPT/GPT-4 was able to recite word by word entire sections of popular copyrighted books, a task that no model is able to do, unless it was trained on those books. OpenAI, even though it was founded as ostensibly an open-source company, has jealously guarded its training data, so it is difficult to tell just how much copyrighted or licensed materials were used as training data without the rightsholders’ consent.

Another example of the harms of secrecy, is Adobe’s Firefly, a model that tries to portray itself as “commercially safe.” Adobe claims that “[Firefly] was trained on Adobe Stock images, openly licensed content and public domain content, where copyright has expired”. It was reported by various media outlets, however, that Adobe Stock Contributors were never given the opportunity to opt-in for their work to be utilized for training, and offers Stock Contributors no way to opt-out their work for use as training data. This imposing denied Adobe Stock Contributors the opportunity to negotiate fair recompense for model ingestion of their work into models that may, and already are, competing against them. Currently Adobe itself has potentially utilized Generative AI works, including my own works, to train Adobe Firefly. Adobe has also refused to be fully transparent concerning the full contents of their training set, prohibiting rights holders from seeing if their works are included, especially concerning the vague category of “openly licensed works”.

It’s important to note that AI companies themselves acknowledge that the use of copyrighted material is an issue for generative AI models. Stability AI for example released the following statement when announcing their music model, Dance Diffusion, in September 23, 2022: “Because diffusion models are prone to memorization and overfitting, releasing a model trained on copyrighted data could potentially result in legal issues.” Unfortunately, while acknowledging the potential for harm in the audio space, Stability did not take the same approach with respect to visual or written arts.

These are only a few examples of how these Generative AI models violate the rights of artists and creatives en masse. The truth I discovered is that current Generative AI models rely on the nonconsensual use of ill-gotten copyrighted data of unwilling artists and the public’s data. This technology is already affecting creative fields like mine, not in a year or two from now, but right now. It is bad enough that this is being done without our consent, without any credit being offered, or without any compensation, but worse still, we are now forced to compete against these Generative AI models that were built upon our own work. No human being can outcompete a Generative AI model, due to the economies of scale: an AI is low cost and can generate a massive volume of “good enough” products especially compared to a single artist.

⁴ This process of using third party organizations to gain access to data companies would normally not get access to has been called by technologist Andy Baio, “Data Laundering.” A lawsuit was recently filed in federal court challenging the use of private information, including personally identifiable information, by OpenAI. *P.M. v. OpenAI LP*, Case No. 3:23-cv-03199 (June 28, 2023), ECF No. 1.

AI companies are able to garner billions in funding and profit on models built from the hard work of artists and creators, while capturing the very market those artists and creators rely on in order to make a living. In any other context, such wide-scale misappropriation of the rights of so many people would be fundamentally unfair.

Therefore, that to reward models that normalize and reward such large-scale theft based generated work that is based on the work of artists and creatives, would be not only deeply unjust, but also immensely damaging to the concept of copyright itself, for why would copyright ever matter if it can simply be ingested and laundered by a model.

II. Generative AI Threatens Jobs

While artists and creatives may be the first to feel the effects of the growth of Generative AI, it is only a matter of time before other professions and industries are affected. According to the International Intellectual Property Alliance's *Copyright Industries in the U.S. Economy: The 2022 Report*, the core copyright industries added \$1.8 trillion dollars of value to the U.S. GDP, accounting for 7.76% of the entire U.S. economy.⁵ According to that same report, “[t]he core copyright industries employed 9.6 million American workers in 2021, accounting for 4.88% of the nation’s workforce and 5.53% of total private employment in the United States,” which is a massive amount of Americans. These risks are not hypothetical, they are already happening. I have personally been a part of three productions where Generative AI has been utilized at different levels. My colleagues at all levels of their careers, both students who are starting out their professional lives or legendary veterans in our entertainment industries have felt the effects of Generative AI. Some of my peers have lost jobs, some have had their duties lessened (and of course, their proposed pay). My entire industry is holding our collective breaths to see how far and how quickly this technology will come to replace us. And it is not only my industry that presently feels the strain, anything that relies on creative effort, from coders to novelists, to filmmakers to voice actors are facing the very real threat of being replaced by Generative AI.

The unfettered growth of generative AI without appropriate guardrails would lead to an unprecedented escalation and use of these technologies in not only creative spaces, but in businesses across all sectors. This could potentially devastate the contributions of workers and the workforce. There are already Generative AI models that code, that write, that generate films, and even generate realistic facsimiles of a person’s voice. It is only a matter of time before more models emerge in other fields that replicate other areas of human expertise. Economists at Goldman Sachs released a report claiming an estimated 300 million full time jobs could be diminished or outright lost from Generative AI. As a point of reference, all full-time jobs in the US and EU amount to approximately 333 million full time jobs according to Statista. The number of jobs threatened by Generative AI amounts to 89.9% of full time jobs altogether in the US and EU. That is a staggering number.⁶

⁵ Available at https://www.wipo.int/export/sites/www/copyright/en/performance/pdf/report_2022_us.pdf.

⁶ The AI companies and AI advocates will be quick to state that there will be enough jobs created in the Generative AI industry to offset these jobs. These claims seem untenable, as for example, Google and Microsoft, two tech giants, employ 356,000 full time jobs worldwide. As generative AI positions itself to diminish or even destroy some of the same jobs of Google and Microsoft employees, I am skeptical of this claim, especially with projected full time job losses in the hundreds of millions.

III. How These Models Work.

Machine Learning (ML) and/or Artificial Intelligence (AI) synthetic media companies, generate and profit from “high quality media” that depends entirely upon training their AI/ML models with datasets. These datasets contain the copyrighted and licensed works as well as the private data of hundreds of thousands of artists, businesses and the general public which have been obtained without consent, credit, or compensation.

AI/ML synthetic media models, let’s say text-to image models, take input from a user in the form of a natural language description, otherwise known as prompts, to generate an image matching that prompt. To condition that capability the model needs to be trained on a huge collection of images, media, and text descriptions scraped from the web and collected in the form of a “dataset ” in order to extract and encode an intricate statistical survey of the dataset’s items. Images are generated from an input prompt by assembling visual data that attempts to best simulate the statistical correlations between text in the dataset and images in the dataset in order to produce “acceptable” results. In other words, generative AI/ML models are completely reliant upon their training and input phase, in order to power the output phase.

These AI companies have frequently claimed that what they are doing is fair, ethical and legal. This claim is not as clear cut as they say. Generative AI models frequently emit output that is identical or nearly identical to material contained in the training data. There are numerous examples from academia. For example, besides the aforementioned study conducted by researched at the University of California, Berkeley revealing near-verbatim reproductions of copyrighted books by ChatGPT/GPT-4, a study by researchers at the University of Maryland and New York University found that generative models generated “high fidelity reproductions” of images found in the LAION dataset at an estimated 1.88%.⁷ Artist Christopher Bretz did the math: “Forgive my napkin math, but taking just Lensa, which uses Stable Diffusion, has ~25m downloads and gives users 50 trial images each. At 1.88% there are potentially 23,500,000 generated images that could be very similar to training data. The portrait targeting might skew that...”⁸ And that number is just for one AI model.

⁷ Gowthami Somepalli et al., *Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models* (Dec. 12, 2022), available at <https://arxiv.org/pdf/2212.03860.pdf>. It is important to note that this particular study is based on a subset comprising less than .06% of training data, and their methods may not be as thorough. This is why the study concludes results systematically underestimate the amount of actual “replications.”

⁸ Christopher Bretz (@saltybretzel), Twitter (Dec. 19, 2022, 6:34 AM), <https://twitter.com/saltybretzel/status/1604847596149047296>.



Steven McCurry's original photograph "Afghan Girl" (left). Generated imagery of "Afghan Girl" via generative model Midjourney (right)



Yuran • Apr 14, 2023
cute puppy in short shorts, humorous, in the style of quentin blake
--s 750

Original Illustration by artist Quentin Blake for "Fantastic Mr Fox" (left) Generated image based on Quentin Blake with prompt for Midjourney Magazine Issue 2 v5 pg 85 (right)

AI companies themselves have warned that their Generational AI models emit identical or near-identical copies of training data. For example, GitHub warns that based on internal research, Copilot, its Generational AI computer code model, "about 1% of the time, a suggestion may contain some code snippets longer than ~150 characters that matches the training set."⁹ Copilot has millions of users. Even 1% of outputs being a near-verbatim match to training data results in at least tens of thousands of results.

⁹ GitHub Copilot, <https://github.com/features/copilot> (last visited July 6, 2023).

And this number is almost certainly an underestimate.

Given that so much of these Generative AI's training data is copyrighted or licensed work that was taken without the creators' consent, it is not difficult to imagine why AI/ML models trained with ill-gotten data are so concerning to creators. Artists and creators who have spent a lifetime honing and refining a skill can now have facsimiles of their hard work reproduced in an instant by a Generative AI model that has been trained on their work without their consent.

Not to mention as these models have shown to be essentially plagiarized content, it could become a legal liability to customers who unwittingly generate identical copies of previously copyrighted work.

IV. Demystifying the Technological Mythology

As policymakers' time and resources are limited, I would like to take a moment and help demystify some of the common mythologies and narratives surrounding Generative AI. Seeing this tech for what it is, and for what it is not, will hopefully illuminate the path forward.

A. False Equivalency: The Anthropomorphic Trap.

A false equivalency is a “is an informal fallacy in which an equivalence is drawn between two subjects based on flawed or false reasoning.” For example, “a blueberry and an orange both taste the same because they are both round shapes” is a false equivalence. False equivalencies are quite frequent in discussions around Generative AI models. In particular, statements that give generative AI anthropomorphic tendencies, or human qualities. It is common to hear “AI/ML models learn just like humans learn” or “artists use references, the machine does the same thing!” But these are false equivalencies—there is an overwhelming consensus of experts in the field that these models do not resemble human beings at all. AI/ML models are mathematical algorithms, which means they are incapable of “learning” as a human being does. A Generative AI model cannot create something new out of whole cloth, it relies on its training data in order to generate its output; it identifies and replicates patterns it sees in the training data. Importantly, it is a human being who programmed the AI to do the things it does. There were human beings who chose what datasets to use as training data, human beings who coded how the AI processes the training data, and human beings who programmed the AI how to convert user prompts into output.

The erroneous anthropomorphization of this technology, however, does serve a very useful purpose. It absolves accountability. It's easy for some to say nothing nefarious is occurring because the fault lies with the “decisions” of the machine. But this ignores the human hands who built said machines and made purposeful decisions to intentionally take and profit from copyrighted work, private data and artists' names without authorization.

B. Humans Are Creative. Machines Are Not.

Stating the obvious again, AI/ML models are not human. AI/ML models can only generate what is instructed to generate and the data it was trained upon. It is completely dependent on the quality imagery of others to generate work.

Artists look at others' work for inspiration. But unlike AI/ML models, while looking at imagery of others is a part of a human's process of learning how to draw or paint, looking at images of others isn't THE singular way we artists learn or create. Artists look at other artists to learn how to solve visual problems and be inspired, but that's as far as looking at other artists will get you. Artists bring their own technical knowledge, problem solving, experience, thoughts and personal lives into each artwork. This is also why humans who are trained in the same way, who paint the same objects will still yield different results. AI's can never accomplish that.

There is also another layer of complexity here. Imitating an artist's work, to the level where it can be difficult to know if the original artist created the imitated work or not, is an extremely difficult and rare skill for humans to acquire. If a human imitates or copies the work of another, depending on what is done with the work of those who successfully imitated others and depending if profit, fraud or identity theft is committed, it could be considered a forgery and against the law. Why would a generated output meant to copy or imitate an artist's work without their consent, or giving any credit or compensation, not also be treated similarly? When humans plagiarize or create forgeries, it is still unethical and illegal. Creators of these machines, or the machines itself should at the very least be held to the same standards humans are held by.

C. Prompts.

At present, according to the Copyright Office, AI-generated work is not copyrightable because in order for work to be copyrightable, it must "owe [its] origin to a human agent."¹⁰ In the quest for synthetic media to gain copyright, we hear that prompts (the descriptions utilized to instruct the models on what to generate) are creative enough to gain copyright. While I am not a copyright lawyer, I would argue, however, that prompts are mere tools or methods of operations for the AI models, and themselves are not enough to be granted copyright.. Further, the Copyright Office has issued guidance that an AI-generated work is not copyrightable when an AI solely receives a prompt from a human because it lacks the "traditional elements of authorship" because those elements are executed by the AI.¹¹ Thus, even if a prompt may be copyrightable, the resulting output itself is not, as currently understood by the Copyright Office.

In addition, even if prompts were themselves copyrightable, I must note there is a push in the Generative AI industry to automate prompts themselves. For example the CEO of Stability AI himself said on Twitter on February 9th, 2023 "You wont need prompts in the future" and again on February 10th, 2023 "Prompt engineering also won't be a thing given you can have AI to dynamically reconstruct and optimize these." This is also evident in recent feature releases by Midjourney that automate various aspects of writing prompts such as their "/describe" and "/shorten" features. So even if prompts were mistakenly considered creative expression, there is a push to incorporate these into the Generative AI models too, demonstrating just how far-reaching these models are.

¹⁰ U.S. Copyright Office, Compendium of U.S. Copyright Office Practices sec. 2.8.3(I)(a)(1)(b) (1st ed. 1973), <https://copyright.gov/history/comp/compendium-one.pdf>.

¹¹ Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 37 C.F.R. § 202 (2023).

D. Generative AI Is Not a Search Engine.

Some would describe Generative AI models similar to search engines. While there are some similarities, they function entirely differently. Search engines rely on indexing and gathering information from vast amounts of data. When a user engages a search engine query of an artist, the search engine will grant users information on the artist, the artist's contacts, and showcase the artist's work. This provides unique economic opportunities for artists as it may expand artist's recognition or publicity, and because users may be introduced to other works the artist may have done. This is true for visual artists, filmmakers, writers, voice actors, musicians, and for other creative persons.

Generative AI on the other hand generates content based on its training data from a user query in the form of a prompt. When an AI user describes an image utilizing an artist's name, Generative AI models completely erase the artist from the equation, by generating imagery or works that are similar to the artist in question for a fee, but without attribution or compensation to the artist. The economic benefits go to the Generative AI companies in this scenario, and not the artist whose work was utilized to generate said synthetic media.

E. Opt-Outs Are Inefficient. Opt-In Is the Only Way Forward.

There is also a push in the generative AI industry to set opt-out (choose not to participate in something) as a standard. I must note that presently, creative persons like myself have never been offered the opportunity to truly opt our work out of training sets, even though this narrative has persisted. Currently, techniques to "unlearn" specific samples of data are extremely costly even for moderate or small models, and completely intractable for state of the art generative AI models. Due to this it is near impossible for algorithms to forget data once it's been trained upon, outside of destroying the algorithms and training sets and starting from scratch.¹² Thus, while Machine "Unlearning" is still a nascent stage, current "opt-out" procedures are largely ineffective and unreliable.

Even if opt-out procedures were perfected, this still misses the mark. Opting out puts the onus on the public to police the use of their data. Given the multitudes of AI models that have proliferated in just one short year, asking a member of the public to constantly monitor the use of their own data by AI companies that are constantly multiplying seems to me to place the burden on the wrong party. For starters, it places an undue burden on people who may not be well versed with the technology, people who may not know the language, people who may have physical impediments, people who may have time limitations, or people who are not on the internet, are unaware their data is in the possession of technology companies and being used for commercial profit.

On a practical level opt-out makes no sense. There is a serious resource imbalance between ordinary citizens and Generative AI companies. Generative AI companies often are well-funded, and generate millions if not billions in profit. Even though they profit richly, Generative AI companies do not want to spend any of those profits in using data responsibly, they instead want to shift the responsibility onto the ordinary citizen. These models are trained upon an immense amount of data, and models are constantly

¹² This particular destructive process is called "Algorithmic Disgorgement" and is a process that the FTC has previously employed in various cases such as Everalbum vs FTC, and as a remedy for Children's Online Privacy Protection Act COPPA.

updated. Does this mean the public has to keep up to date every time a new update to a model comes out? What about every time you publish something new online, does one have to verify if it is used by Generative AI companies? What about when third party users upload data without your consent? How does one keep track if their work was utilized? How does one find said work in the millions of billions of data points? What if you miss the opt-out period, and then become an unwilling participant? Technology companies cannot claim ownership of all the data on the internet that exists for our benefit without paying heed to copyright or privacy rights. Opt-out would be an ineffective and inappropriate standard for most if not all consumer facing industries, Generative AI should not be the exception. The onus should be on the technology companies who devote their immense wealth to build these models to ensure they are doing it fairly and ethically, not on the common citizen who may not even be aware their data has been taken. Explicit opt-in is the only way forward, as it helps ensure that all AI/ML models are built with right holders and the public's explicit authorization for commercial use of their data from the beginning.

V. Potential overall solutions.

Again, I must reiterate that I am not opposed to AI. But I am interested that AI develops in a way that is fair, ethical and equitable to all participants, whether it is the technology company, the user, or the person whose data is being used to train the AI. Below are some suggestions Congress can implement in order to ensure that Generative AI truly serves the public, without exploiting the rights and livelihoods of those whose data feeds the AI:

1. Congress should exercise its constitutional authority¹³ to update the Copyright Act to reaffirm that copyright requires human authorship.
2. Congress should empower existing agencies to regulate the use of data to train Generative AI's. 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. At all steps Congress should take care to avoid regulatory capture to ensure stakeholders have a voice in the regulatory process. This can take the form of ensuring technical and academic experts who have not been employed by AI companies have a role in determining 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 have a role in determining regulatory policy. Different considerations should also be given to different industries, as what could be beneficial to the healthcare sector, could be devastating to the entertainment sector.
3. 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

¹³ U.S. Const. art. 1, § 8, cl. 8.

permission. This can also take the form of passing a law authorizing a federal civil right of publicity cause of action.¹⁴

VI. Closing

I love what I do. Deeply. My art is something I have spent my entire life working towards developing. I love my craft and I love to teach those who love it just as deeply as I do. It is part of who I am. And I know that sentiment is shared by all of the artists, writers, coders and others who have had their work ingested and trained on by Generative AI without their consent. It is because of that love that I am here today. Ultimately, you as congress are faced with a question about what is fundamentally fair in American society. Is it fair for technology companies to take work that is the product of a lifetime of devotion and labor, even utilize creators' full names, without any permission, credit or compensation to the creator, in order to create a software that mimic's their work? Is it fair for technology companies to directly compete with those creators who supplied the raw material from which their AI's are built? Is it fair for these technology companies to reap billions of dollars from models that are powered by the work of these creators, while at the same time lessening or even destroying current and future economic and labor prospects of creators? I'd answer no to all of these questions.

My livelihood is threatened as a result of the uninhibited growth of Generative AI. And I am not alone. Indeed, I and artists like me may only be the first wave of Americans who will have their livelihoods erased by the onset of Generative AI. But tomorrow it could be any number of Americans in a multitude of other professions who may be replaced.

This is why I implore the esteemed members of this subcommittee and Congress as a whole, to truly consider the implications of what it would mean to not only the millions of creative professionals that are threatened now, but the professions next in line under threat by AI if nothing is to change. What it means when unethical machines are rewarded with the same legal rights as humans, in the form of copyrights, to replace the creative soul of a nation. And what it means for potentially unprecedented levels of jobs lost because of this same technology.

I implore you all to please consider next steps carefully, to ignore the hype, and understand that progress can be attained in a way that does not sacrifice whole industries and the millions of jobs that power those industries. If there is one last thought I can leave you, is that this is not zero-sum game. We can have the benefits of these technologies, while respecting the copyright and privacy of rights of those whose data is being used for training. And you as Congress have the power to shape law and policy by saying that the property rights and copyrights do matter. That privacy rights do matter. That American livelihoods do matter. And that just because a technology is new does not mean that it excuses the wide scale infringement of the rights of creators and the public. Thank you.

¹⁴ While under the laws of many states, a civil plaintiff may bring a case asserting violations of the right of publicity, it is often the case that many of those cases face difficulties due to preemption based on the Copyright Act or under state anti-SLAPP laws. A federal law would eliminate many of those hurdles.