Copyright Office Inquiry on Artificial Intelligence and Copyright
Comment by Chelle E.G.

Below is an account of the current harms generative AI poses to myself and my peers in the arts through the non consensual appropriation of our works and likenesses. Following this introduction are my answers to many of the Copyright Office's questions with citations and a list of supporting papers and articles.

My Experiences and Concerns:

Even before the proliferation of generative AI, I have experienced loss of income and wellbeing from my copyrighted works being appropriated by individuals and companies from around the world.

I have been drawing since childhood, dedicating over two decades of my life to developing the motor skills, experience, technical knowledge, and memory required to illustrate fanciful characters and creatures from my imagination with both traditional and digital painting mediums. For artists like myself, there is little that is more rewarding than starting with a blank canvas and bringing a story or vision to life with graphite, ink, paint, or pixels. Despite pressures by schooling and society to pursue careers in other fields, myself and many of my peers in art and entertainment continue to endeavor in our craft because it is so fulfilling and personal to each and every one of us.

The personal nature of the arts is in part why the violation of our rights, including copyright, hurts so deeply. Years ago I painted my own rendition of an ancient mythological creature. At the time, there were relatively few modern illustrations of this creature (when compared with dragons, for example), so when I published my piece to my portfolio it was met with praise and even became one of the first results in search engines like Google. At first I thought that my piece's popularity online would attract clients and buyers for my work- but my excitement was swiftly stifled by the nigh-unending infringement that followed.

Over the past five years I have issued over a thousand DMCA notices against sites and businesses using my art in their advertising, apps, webpages, videos, blog posts, 'free' image repositories, logos, NFT collections, and even offered for sale on physical merchandise. I have spent hundreds of hours scouring search engines, submitting report forms, and drafting emails just to assert my rights and stop others from illegally profiting from my work. Never once have I been compensated by infringers or the platforms that host them (among those platforms are Youtube, Amazon, eBay, Opensea, Etsy, X/Twitter, Tiktok, Tumblr, Pinterest, Meta's platforms Facebook and Instagram, and countless more), even in instances where they profited from sales of my infringed works.

It became frustratingly apparent that the Digital Millennium Copyright Act of 1998 did not anticipate the extent to which infringements would occur online. That the 'safe harbor' clauses provided by Section 230 would heavily disincentivize any preemptive moderation on the part of platforms- leaving the burden of policing millions of websites to already overworked and independent creative professionals.

Ultimately, I decided it was necessary for my health to either make my online portfolios private or delete them entirely, which greatly reduced the number of infringing uses of my work. However, the act forced me to put my career as an artist on hold to pursue other creative occupations- like voice acting, and joining my husband as a reviewer and online creator in the video game and tabletop gaming industries. I am fortunate in that I had these alternatives, but

many artists do not have the support system or ability to stop marketing their work online, even if only temporarily. Regardless of the circumstances, no artist should be forced to abandon the careers they worked so hard to achieve because of the misuse and exploitation of their own copyrighted works.

After combatting infringement for years, it was no surprise, then, to discover in 2022 that the LAION-5B dataset (derived from Common Crawl indexes of images on the internet) included hundreds of copies of illustrations my husband and I have painted over the past ten years. All collected and used, again, without our consent. This dataset is where many AI image generators source the 'training data' required to build image generating AI models.

Upon seeing my works harvested for the purpose of 'training' AI, I again felt violated and taken advantage of- but in a manner I do not know can ever be rectified. My paintings and performances are the culmination of my experiences, imagination, and practiced skill developed over the course of a lifetime. For my work to be reduced to 'data', to be forcibly incorporated into products that stand to make companies billions with seemingly no recourse, has made it difficult for me over the past year to find any joy in the act of creation. Why share what I create with others if it will taken and used to displace and devalue myself and my peers? It has left me feeling hollow, directionless, and fearful for my future.

Now, when I look for images of the aforementioned mythical creature I painted in search engines, the majority of results are obviously AI generated. The very same unscrupulous sites and sellers that once infringed upon my work to make their products almost exclusively use AI generated images instead. While these predatory businesses no longer directly take and profit from my artwork, it is likely that my piece is one of those used by the AI generators they employ to formulate derivatives. In this respect, they are still unfairly profiting from my labor, and that of every artist whose works were used to make their AI outputs possible.

It had been difficult before the proliferation of generative AI to convince clients and companies to fairly pay artists for their labor through licensing, royalties, or even work for hire- but in the past year I have seen accounts from my peers, many who make a living off of commissions or freelance work, that they have seen their former employers stop hiring artists entirely and turn to AI generators for their imagery and assets instead.

Assertions of 'fair use' fail when the outputs of generative AI products directly compete with the copyrighted works they use as 'training data'.

And beyond illustration, my peers in the gaming industry (performers, influencers, actors and voice actors), are finding that their likenesses and voices are being turned into custom AI models to mimic them- sometimes maliciously. SAG AFTRA (the Screen Actors Guild) joined the WGA (the Writers Guild of America) in striking against unfair contracts in the television and film industry in part because of concerns that actors' likenesses and voices will be used in AI models to replace them. However, 'deepfakes' are a real danger to any US citizen who has ever been photographed or recorded. Many AI models, especially editable 'open source' models, allow users to directly input images and recordings to manipulate them. With this, bad actors are capable of creating pornography of celebrities and minors, potentially swaying elections or court cases by falsifying events and evidence, scamming people over the phone with the voices of their relatives, or tainting the information ecosystem with plausible sounding text or realistic imagery.

I live presently in a constant state of fear and dread. Fearful that my voice or visage will be used against my will to say heinous things or partake in acts I would never condone. That my art will be consumed to generate outputs that will compete with myself and my peers in the market. That platforms I have used for the better part of a decade will use my works and

private data. That perhaps, in the near future, I may not even be able to distinguish fact from fiction.

The harms to myself, my peers in creative industries, and to the general public are too great to ignore. Protections for creative works, as well as protections to likeness for every US citizen, must be expanded and enforced in light of generative AI. The 'safe harbor' loophole of the DMCA, which places undue burdens on individuals to police for rampant infringement, should not be repeated with the unfair expectation that US citizens must scour datasets and 'opt-out' of generative AI models. There must be swift and stringent consequences for the violation of rights- as too often companies, platforms, and infringers have skirted responsibility without being required to compensate those they harmed.

Artists were already devalued and struggling to protect their works before, and AI presents yet another means for their labor to be stolen and appropriated. The exploitation of US Citizens' labor and likenesses should never be condoned in the name of technological progress. I have written to senators' offices, and one (Senator Chuck Schumer) sent a response indicating that the government is seeking a 'balance' between human creators and technological innovation. However, innovation should never be at the expense or detriment of US citizens and their rights. There can be no balance if artists and US citizens are deprived of the right to refuse the use of their works and likenesses as grist for machines.

The typewriter did not appropriate the works of writers. The camera did not appropriate the works of painters. The mechanical loom was not constructed from the flesh and bone of textile workers. Generative AI models, however, require the faces, voices, music, prose, artworks, and expression of creatives and performers to generate what makes them valuable. That value, estimated in the billions. must be distributed amongst those who contributed to it, or the whole system will be built off of theft, exploitation, and infringement.

I must stress that copyright is so important because artists and other creatives stand at the unique intersection between labor and the owning class- where they are manual laborers in their craft who then own valuable intellectual property that they can use to make a living where they are afforded freedom and independence.

Generative AI, thus, threatens to devalue artists' labor and transfer the value of that labor to AI developers, powerful corporations, and unskilled competitors. This exploitative extraction of value from independent laborers by companies that do not compensate them for the value of their work must be made clearly unlawful if creative professions, and perhaps even copyright itself, are to continue to exist.

Thank you so much again for reading my personal account and thoughts.

-

I suggest reading "AI Art and its Impact on Artists" by AI researchers and artists as well for their observations on the chilling effect of AI on genuine human artists and creators. [1]

[1] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. "AI Art and its Impact on Artists". In *AAAI/ACM Conference on AI, Ethics, in Society (AIES'23), August 08--10, 2023, Montreal, QC, Canada.* AMN New York, NY, USA 12 Pages. doi.org/10.1145/3600211.3604681

Answers to the Questions Posed by the Copyright Office:

1 and 2:

I work within entertainment industries that rely on creators contributing their performances, voices, music, and art to film, games, live streaming, literary publications, and more. The proliferation of AI generated material has already threatened our livelihoods and rights.

-Impersonation and Appropriation of Likeness: AI 'Deepfakes' allow unaffiliated parties and bad actors to take the faces and voices of performers and celebrities to output non consensual pornography, falsified recordings, or even program models that serve as personal 'chatbot companions'.

As a performer and voice actor, I have been paid to provide my voice for video game characters, trailers, and commentary. Actors in the video game industry are particularly vulnerable to replacement by AI, as game developers often forgo voice acting in their games to save in their budgets. AI voice cloning would allow studios to cut voice actors out of the industry entirely, and potentially with the very voices of those displaced actors if laws are not put in place to discourage the practice. Actors are also concerned that past projects may use vague contract clauses to secure the rights to use their voice 'in perpetuity'.

It is thus concerning that third parties using AI could 'clone' our voices and puppeteer duplicates of our bodies while profiting from features unique to us. Voice and likeness, while marketable offerings for actors, are also integral parts of their identity. Any AI use of a person's voice or likeness should be considered a violation of their rights.

Examples of notable deepfakes in the industry include recently disseminated advertisements for a giveaway featuring Youtuber MrBeast [2], and deepfake porn of Twitch streamers like Pokimane, QTCinderella, Sweet Anita, and more. [3]

[2] Whateley, Dan. "MrBeast Calls Out Tiktok for Allowing a Deepfake Version of Him in Ad" *Business Insider.* October 3, 2023
www.businessinsider.com/tiktok-ran-deepfake-ad-mrbeast-as-ai-generated-content-spreads-2023-10

[3] Britton, Bianca. "They Appeared in Deepfake Porn Videos Without their Consent. Few Laws Protect Them." *NBC News.* February 14, 2023.
www.nbcnews.com/tech/internet/deepfake-twitch-porn-atrioc-qtcinderella-maya-higa-pokimane-rcna69372

-Unfair Competition and Oversaturation: Those who use generative AI are able to swiftly output content that floods the marketplace and competes with completely human authored works and performances.

For instance, depending on the complexity of an illustration, I find that it can take anywhere from a few hours to several days of work to craft a finished painting. These estimates do not take into account the years, if not decades, of practice necessary to develop the skill required to paint at all, much less at the quality sought by clients and buyers. An individual using generative AI, however, could output dozens of 'high quality' images in a matter of minutes with no prior artistic capabilities or knowledge. These artificial outputs can then be uploaded en masse to marketplaces and social media to compete with genuine authors artists on sites like Amazon [4] and Etsy [5]. I have even witnessed AI users bringing tangible prints of their

outputs to craft fairs to compete with the handmade paintings of seasoned artists. A human creator cannot physically paint, type, or craft swiftly enough to produce at the same pace as automated AI systems.

It is also possible that the presence of AI generated media will far outstrip the cumulative creative works of human artists within the next few years, particularly in online spaces that used to serve as marketing outlets for creators. Even if artists' works are properly licensed and paid for by AI developers, the outputs of generative AI would still overrun marketplaces and taint the information ecosystem with artificial art, photo-realistic images, and convincing audio and text.

[4] Duffy, Clare. "An Author Says AI is 'Writing' Unauthorized Books Being Sold Under Her Name on Amazon"- *CNN.* Aug 10, 2023
www.cnn.com/2023/08/10/tech/ai-generated-books-amazon/

[5] Tiffany, Kaitlyn. "AI-Generated Junk Is Flooding Etsy"- *The Atlantic.* June 15, 2023
www.theatlantic.com/technology/archive/2023/06/ai-chatgpt-side-hustle/674415/

-The Devaluation of Skilled Work and Wages: Just as outsourcing and automation destroyed domestic manufacturing jobs and lowered pay for workers in several US industries, generative AI is already replacing skilled creatives in their respective fields. Unlike prior technological innovations, however, generative AI relies on its consumption of the labor of human creators to attain its value.

My peers in illustration, unfortunately, have already relayed stories on social media of the loss of once consistent clients and licensing deals as AI generators are used to output artificial illustrations for books, games, and merchandise. The irony is that many of these generative AI models are only capable of outputting similar works because they utilized the existing art of illustrators without their knowledge or consent.

Many AI companies offer their models' services for free, or with purchased 'credits' or monthly subscription fees. These fees, if present, are typically around 5-20 USD per month for the generation and purchase of potentially hundreds of images. An artist, even if working at the US minimum hourly wage, could never charge competitively against such low subscription prices.

Human creators cannot feasibly compete with cheap and free AI image generators, especially when the generators were built by consuming creators' prior works.

-The Loss of Multidisciplinary Collaboration: Especially in film and games, collaboration between creatives is what makes them so varied and vibrant. Writers, directors, artists, actors, musicians, 3D modelers, costumers, and more are brought together to contribute to projects that provide them not only with employment, but community.

I have in the past year, unfortunately, been approached numerous times for coverage by video game developers who used generative AI for in-game assets and marketing images. These developers often claim that they do not have the budget to hire an artist, and so resort to using AI models that likely appropriated the works of those artists. Some have also stated that they do not want the hassle of corresponding and working with others, and would rather use AI to generate images, voices, and music that suits their needs. It has been disheartening to see creatives of other disciplines turn on artists by using generative AI to replace the need for them to work with one another.

This is in part why the WGA and SAG AFTRA strikes began- because even multibillion dollar corporations refused to negotiate on proposals that would prohibit them from freely using the writing, voices, or likenesses of their workers in generative AI for film and television production. [6]

If creators do not have the ability to prevent their works and likenesses from being exploited by AI, and cheap AI alternatives using their works and likenesses are offered to individuals, studios, and corporations- those creators and performers may be entirely replaced in industries that once thrived because of collaboration.

[6] Peterson, Andrea. "SAG-AFTRA is fighting a dystopian AI takeover so you don't have to." *MSNBC.* July 18, 2023.
www.msnbc.com/opinion/msnbc-opinion/wga-sag-aftra-strike-writers-actors-ai-hollywood-rcna94318

-Fear of Celebrity and Exposure: Unfortunately, I have witnessed professionals and hobbyist artists, voice actors, and writers alike reveal instances in which their works, faces, and voices are utilized by AI companies and consumers without consent. Creators who are well established are often the first to be victimized because of their fame and the proliferation of their works across the internet. Even in the web scraping utilized by AI developers, the more prolific the artist, the more likely their works are to be scraped and copied several times over in datasets. Understandably, famous artists' names are being used in 'prompts' to mimic their works in AI outputs.

And, alarmingly, when artists voice their concerns and request their works be removed from models or that they and other artists deserve to be compensated, they are met with vitriol and abuse. [7] AI developers, dataset makers, and AI users frequently attack artists on sites like X/Twitter, Reddit, and more. They argue that artists are at fault for uploading their art online because AI advocates believe 'everything public is free to use'. AI users often try to assert that 'copyright should not exist' and that it does not protect artists works from being ingested by AI. They frequently taunt artists by saying that they need to 'adapt or die'. And, if artists speak out against generative AI, that their art will be intentionally fed into AI models to produce outputs that mimic their work. For artists and other creatives, it is horrifying to see their works taken and altered by those who vocally state that they want to make them 'obsolete'.

I made my own social media accounts private in the years leading up to this current wave of harassment, in fact, because NFT proponents similarly targeted and infringed on the works of artists by selling their works on NFT and Crypto marketplaces. Myself and many of my peers are now afraid of using social media at all because of the abuse levied at artists and the ease with which AI users can appropriate our works with 'fine tuning' and 'image to image' applications. This is also why I am not divulging my full name in this public document and comment lest I be targeted as well.

[7] Nuttall, Jeremy. "Whose art is this, really? Inside Canadian artists' fight against AI". *Toronto Star.* February 2, 2023.
www.thestar.com/news/canada/whose-art-is-this-really-inside-canadian-artists-fight-against-ai/article_54b0cb5c-7d67-5663-a46a-650b462da1ad.html

To creators like myself, it feels as though those who benefit most from generative AI are largely corporations and individuals who want to take advantage of creative markets with minimal effort or use AI as a means to phase out hiring creators or the need to purchase or license creators' works. Despite the potential 'benefits', the way in which current generative AI models are constructed with unlicensed copyrighted works and likenesses taints the whole system.

If the United States and Copyright Office assert that all 'training data' must be licensed and the consent of copyright holders sought prior to the use of their work, it would do much to stop the cycle of abuse artists are currently enduring.

4.

International consistency in regulating AI will be necessary to protect the rights of US citizens' across the globe.

Examples of foreign regulatory measures being considered or implemented in regards to AI include:

-The EU's upcoming 'AI Act' demands data transparency for the materials used to train AI models, particularly a declaration of the use of copyrighted works. [8] AI developers and companies would be obligated to submit frequent reports on data sources to the European Commission, in part with the intention of helping copyright holders discover instances where their works were used without permission. Additionally, if AI outputs cause harm (like helping users build a bomb) AI companies and developers could be fined 6% of their yearly revenue or possibly be banned from operating in the EU. Electricity usage would also need to be reported, as the computing power required to run and train AI models has more than tripled its share of power consumption in countries like Ireland. [9]

[8] Gordon, Anna and Perrigo, Billy. "E.U Takes a Step Closer to Passing the World's Most Comprehensive AI Regulation." *TIME.* July 14, 2023.
time.com/6287136/eu-ai-regulation/

[9] O'Carroll, Lisa. "EU 'in touching distance' of world's first laws regulating artificial intelligence". *The Guardian.* October 24, 2023.
www.theguardian.com/technology/2023/oct/24/eu-touching-distance-world-first-law-regulating-artificial-intelligence-dragos-tudorache

-The United Kingdom struck down clauses in its own upcoming AI policies that would have allowed AI companies to train off of unlicensed copyrighted works for commercial purposes. They did not implement the exemption because it would 'risk reducing arts and cultural production as mere 'inputs' in AI development'. [10]

[10] Creamer, Ella. "MPs criticise UK government's handling of copyright policy related to AI". *The Guardian.* August 30, 2023.
www.theguardian.com/books/2023/aug/30/mps-criticise-uk-governments-handling-of-copyright-policy-related-to-ai

-China has drafted regulations that would demand that AI generated outputs be labeled as such with watermarks, and bars the creation of deepfakes without the consent of the subject. They are also considering laws that would prohibit private data and intellectual property from being used in training AI products. [11]

[11] Yang, Zeyi. "China isn't waiting to set down rules on generative AI". *MIT Technology Review.* May 31, 2023.
www.technologyreview.com/2023/05/31/1073743/china-generative-ai-quick-regulation/

We must consider these and other international regulations in the development of our own- as the copyrighted works, likenesses, and private data of US citizens are being harvested by AI developers around the world.

5.

Existing Copyright law should be clarified and/or expanded to officially recognize that the unlicensed use of copyrighted works as 'data' in generative AI development is a form of infringement- especially when the works are used to create generative AI models whose outputs can compete in the market with the original works. Just as an artist should have every right to choose what publishers, individuals, organizations, businesses, or platforms have lawful license to profit from their works, so should every US citizen have the right to demand that AI developers seek permissions for and license their likeness, personal data, and intellectual property as well. It should be a crime to input unlicensed works into datasets and AI models- resulting in severe statutory fines, renumeration to impacted parties, and the injunction, disgorgement, and destruction of infringing AI models and datasets.

I believe that the Copyright Office is correct in its ruling that wholly AI generated media would not be afforded copyright protection. I argue, however, that while the Office is considering allowing for partially AI generated media with 'substantial human authorship' to attain a form of copyright, that no AI generation should be granted copyright protection if the underlying AI model used unlicensed works in its training data or other inputs. The models used and any additional inputs should always be divulged in any attempt to copyright works that employed generative AI, and the Office should be aware of what models use unlicensed works and are therefore infringing.

No AI generated media should be afforded protection if it was even partially made with unethical and infringing models that profit from the exploitation of artists' and US citizens' works and data.

It will likely be challenging for The Copyright Office to define what elements of a partially AI generated work (using only properly licensed models) would be eligible for copyright. There are AI users who employ generators to make an underlying image that they then edit, for instance. Oftentimes, however, the final image may only look like a more refined version of the original generated image. I do not believe such an image should be eligible for copyright protection, as the AI constructed its integral elements.

Subject, composition, anatomy, color, lighting, detailed objects, abstract forms, medium, line quality- an artist makes conscious decisions at every step in the construction and execution of their manual drawings and paintings. An AI user would be hard-pressed to demonstrate that they have the control necessary to impact every one of these features in AI outputs. Even should they use collage to input an image that the AI then modifies to make it more cohesive and refined, (giving the user more control of the underlying composition)- the AI would still give the piece its final quality, veneer, and appearance. The AI, not the user, would be responsible for the final execution.

An AI user, ultimately, is still utilizing a randomizer to generate assets. As such, it is difficult, in my mind, to grant even 'partially' AI generated works copyright protection if the user cedes the creative process in composition and/or rendering its final aesthetics to a machine built with the creative expression of others.

New legislation will most definitely be required for data privacy rights and publicity rights so that platforms, companies, and individuals cannot non-consensually use and profit from the data, voices, likenesses, and biometrics of US citizens.

Suggestions for other proposals: Many of these are related to how AI developers attain copyrighted works and private data and well as the publication of AI generated media. These may not fall under the Copyright Office's purview, and may be handled by other government organizations, like the FTC. Additional proposals will be under relevant questions.

-Forbid operating systems, email services, cloud storage, software, social media platforms, websites, and marketplaces from assembling datasets and training AI models with the text, images, video, and audio created and uploaded by their users. For example, Meta, the owners of Facebook and Instagram, recently released AI models trained on posts by their users. [12] Permissive and vague Terms of Service and Privacy Policies should not be used as a way to acquire licenses from users to their works or sub-license their works to third parties. This is especially necessary as creators and the general public rely on services like social media to communicate with their friends and family, network, and market themselves and their copyrighted works online. Private emails, messages, and video calls should never be used as 'training data'.

Additionally, as social media platforms and and other repositories of media frequently host infringing content uploaded by third parties, they should not make blanket assumptions that their users have the authority and the right to license media for AI training. As mentioned prior, I have issued over a thousand DMCA notices against infringing uploads of my work found on various social media sites. It is, therefore, likely that my artwork has been ingested into AI datasets via infringing posts and copies.

Only sites specifically designed with the narrow focus of licensing (like stock image and licensing websites for other media) should offer licenses to AI developers. Creators who already use these sites for conventional licenses should have a say in whether or not they want to license their works for AI development as opposed to clients that license their existing work as-is.

[12] Paul, Katie. "Meta's new AI assistant trained on public Facebook and Instagram posts". *Reuters.* September 28, 2023. www.reuters.com/technology/metas-new-ai-chatbot-trained-public-facebook-instagram-posts-2023-09-28/

-Ban web scraping for the purpose of compiling datasets and training generative AI. Much of the media online is copyrighted or infringing copies of copyrighted works, and therefore should not be collected en masse for use in AI development.

-Laws may have to be reconsidered revolving around copyright and licensing images and recordings that depict the face or voice of another individual that is not the copyright holder's. People out in public may unwittingly be photographed or recorded by other individuals who can then effectively license their likeness to AI developers because the recorder owns copyright to the images. It may be necessary to assert that every individual bears a right that likens to copyright to their own face and voice. That would allow them to prohibit others from publishing media depicting their likeness to sites that develop AI or license media with their likeness to AI companies.

For example- at a theoretical family gathering an aunt could innocently take a picture of her niece and then upload the picture to a social media site that develops AI with any images uploaded to it by its users. I believe the niece has the right to request the removal of the picture, their image from the social media's training datasets, and the destruction of the AI model that ingested their visage. The aunt may have had the niece's consent to take her

picture for private records and personal use (like printing it for a framed photo to place in her home or within a physical photo album) but not to give license to the niece's likeness to an AI company.

This may have implications for news organizations, paparazzi, and those who take and publish images and recordings of individuals without their consent. While they may be able to take and publish the photos and recordings, they should not be able to license and sub-license them for AI development. New privacy laws will be necessary with AI models' capacity to appropriate and manipulate the images of the public.

-Smart devices, like modern TVs, phones, security cameras, computer 'assistant' products (like Alexa or Dot), and more should not be used to harvest information for training AI. This is because these devices can capture conversations and visuals from environments populated by minors and people who are potentially unaware of their presence and cannot consent to being recorded. US citizens may not be aware as to the extent these devices record their conversations and activities in the privacy of their home or in public spaces, and should not have their privacy invaded by surveilling devices.

-The United States should consider laws similar to the GDPR and 'right to be forgotten' that would allow individuals to remove their data from websites and AI models.

-AI generated media that used models with unlicensed works (billions of images may have already been uploaded to the internet in the past year alone) should be reportable. Media made with these infringing models should be removed from public display, sale, and distribution, so as to mitigate market harm to affected copyright holders.

-Generative AI should be banned from autonomously uploading generated media and making fake interactions on social platforms. The potential for bad actors to use generative AI to make bots outputting convincing text, images, and profiles that can skew discourse and be implemented in catfishing and other scams are too great. This is especially necessary to protect children from being manipulated online.

Chatbots may have a place in customer service, but customers should always be given the option to speak to a human representative. Laws should be put in place that require businesses to employ enough human customer service members to meet the demands of their customers for 'human consideration'.

-Generative AI should be prohibited from being employed in writing articles, news reporting, textbooks, scientific papers, encyclopedias, and text and image libraries that are reserved for factual and historical accounts and reference. It will be imperative that there are trusted resources accessible to the public (via libraries and vetted sites) that do not contain any AI generated media for historical preservation and research. False AI fabrications, referred to as 'hallucinations' by AI developers, should not be possible in any model intended to provide facts to consumers, including search engines.

Because of generative AI's capacity to output false information, it should not be trusted in news and education. Such AI models are a danger to the public, as witnessed in AI generated cookbooks, mushroom foraging books published on Amazon [13], and cleaning instructions that promoted the consumption and use of toxic substances.

[13] Milmo, Dan. "Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI". *The Guardian.* September 1, 2023. www.theguardian.com/technology/2023/sep/01/mushroom-pickers-urged-to-avoid-foraging-books-on-amazon-that-appear-to-be-written-by-ai?ref=distilleddaily.news

-Require that medical facilities and businesses who use scanning technology to collect biometric data attain affirmative consent and explain how the data is being used and where it is being stored prior. Myself and my family members (years ago, long before we were aware of how this technology could be abused) had allowed our feet to be scanned at stores like 'Fleet Feet' for the purpose of finding shoes that would be more comfortable for our foot shape- but unwittingly may have licensed our foot scans to their partner 'Volumental'. We were under the impression from the store's brief explanation that our foot scans would be used in a closed system to compare our foot shape to the shoes being offered- not to be sent back to a data company for other uses.

Medical patients should also be allowed to refuse photos and scans of their bodies for use in AI development, even if they are promised that the data will be made anonymous. This should be considered a violation similar to HIPPA if the patients do not give their full consent.

Cameras in public places, like security cameras and cameras mounted on vehicles, should not be used to collect biometric data for AI as well.

Training:

6.

Copyrighted materials, in the form of text (books, articles, messages, social media comments), images (art, illustrations, photographs), video (films, television series, news broadcasts, social media shorts), audio (music, voice recordings, video audio, audio books, podcasts, phone calls) and more are used to train generative AI. Some may be licensed (with websites starting to force the licensing of their current users' media), but many that are unlicensed are scraped from web crawlers and search indexes or downloaded from websites and piracy platforms. [14] If there is curation, it is sometimes outsourced to firms overseas in Africa and Asia, where underpaid workers sift through the 'data' for labeling and to moderate CSAM and violent, illegal, and disturbing media. [15] The practice of web scraping, understandably, raises concerns on privacy as well as copyright.

[14] Lazarro, Sage. "The A.I. industry confronts life after data scraping". *Fortune.* August 29, 2023. fortune.com/2023/08/29/the-a-i-industry-confronts-life-after-data-scraping/

[15] Rowe, Niamh. "'It's destroyed me completely': Kenyan moderators decry toll of training of AI models". *The Guardian*. August 2, 2023. www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

6.1:

Some developers of AI models openly admit that they acquire copyrighted works from web scraping or downloading the contents of search engine indexes, social media, and potentially thousands of websites. These works are collected into datasets that can then be distributed through repositories like Github, Hugging Face, and others. The contributors to these

repositories include individuals, researchers, and private companies who invest in the creation of datasets. One dataset, called LAION-5B, was created by a German team under the guise of research, but with investments from Stability AI (the UK developers of image generator Stable Diffusion)- which fuels many for-profit downstream AI apps like Lensa that rely on the unlicensed and uncompensated use of copyrighted materials. [16] The Books3 dataset, comprised of over 180,000 pirated books, has allegedly been used to train text generators by Bloomberg and Meta. [17] [18]

I have confirmed that my own artwork and photographs are present in the LAION dataset, through sites like 'HaveIBeenTrained' by Spawning that display the contents of the dataset through linked embeddings. Many AI models do not disclose where their data comes from in any capacity.

With the accessibility of 'open source' models, there are communities of disparate individuals that are able to download copyrighted works from the web to input them into custom datasets and 'fine tuned' models.

Additionally, companies like Alphabet/Google, Meta/Facebook, X/Twitter, and Microsoft have modified their Terms of Service with language that would allow for them to potentially train AI models off of their users' personal files, social media posts, biometrics, and more. I have personally reviewed some of these these new terms, and am disturbed at the implication that my private photographs, messages, and other personal data are being used as well as my copyrighted works. It is possible through these terms that the companies will then try to sub license and sell that data to other AI developers.

In places that formerly championed authenticity and creativity, artists like myself are swiftly seeing the promise and opportunity the internet once provided for them being used to exploit them. It is apparent that the rights of US citizens to their creative works and personal data are being violated by the deliberate scraping of media from the web and new licensing terms being added to popular digital services and platforms.

Transparency of 'training data' must be mandated to ensure that works are not being used without proper license and consent.

[16] Javaid, Maham. "The magic Avatars you paid $3.99 for is probably stolen, artists say." The *Washington Post.* December 9, 2022.
www.washingtonpost.com/technology/2022/12/09/lensa-apps-magic-avatars-ai-stolen-data-compromised-ethics/

[17] Leffer, Lauren. "Your Personal Information is Probably Being Used to Train Generative AI Models". *The Scientific American*. October 19, 2023.
www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/

[18] Asmelash, Leah. "These books are being used to train A.I. No one told the authors." *CNN.* October 8, 2023.
edition.cnn.com/2023/10/08/style/ai-books3-authors-nora-roberts-cec/

6.2:

As mentioned in 6.1, websites and services investing in their own generative AI or with an incentive to give their users' data to other AI developers will reinterpret their Terms of Service and Privacy Policies to allow for extensive AI licensing.

One such company is Adobe, which trained its own generative AI 'Adobe Firefly' on the contents of its platform 'Adobe Stock'. Adobe Stock artists and photographers have publicly stated that they were not given the opportunity to 'opt out' of licensing their works for use in Firefly. [19] Adobe has since lauded the generation of over 3 billion images with Firefly, when the corpus of their original stock offerings from artists and photographers only numbered in the millions. Thus, in a matter of months, AI generated media has vastly overtaken the human-made art and photography available for licensing on the Adobe Stock platform.

Where once contributors were compensated by directly licensing their images through Adobe Stock, it is now possible that former stock buyers will eschew purchasing licenses to existing works in favor of generating images with Firefly. It was reported that 90% of Firefly users had never used Adobe products prior. [20] This means that it is possible the majority of Firefly users were not digital artists or photographers prior to the release of Firefly.

There are also disturbing accusations by artists posting on X/Twitter that unlicensed images or AI generated images mimicking their copyrighted works (generated by using AI models that likely had access to their works and produced images similar to their own with prompts including their names) have been uploaded to Adobe Stock. [21] This nullifies the credibility of any licensing scheme for AI training material that is dependent on unverified uploaded assets to a platform. This will be especially true for social media platforms, where infringement is rampant and unlicensed media is distributed freely.

In light of Adobe's pursuit of AI in spite of the harms it presents to artists, my husband and I have unsubscribed from their digital illustration and video editing products- products that we paid thousands of dollars to use and had been working with professionally for over a decade.

As referenced in prior questions, some companies claim they have a license to the media they use in their models; but it seems many rely on reinterpreting their social media or marketplaces' terms of service to give them permissions to the works of their users for AI training with no option to 'opt in' or 'opt out'. This practice allows platforms to unfairly exploit artists and extract value from their work while displacing them in the market, and should be discouraged if not outright banned to protect US citizens and copyright holders.

[19] Goldman, Sharon. "Adobe Stock creators aren't happy with Firefly, the company's 'commercially safe' gen AI tool". *VentureBeat.* June 20, 2023.
venturebeat.com/ai/adobe-stock-creators-arent-happy-with-firefly-the-companys-commercially-safe-gen-ai-tool/

[20] Lardinois, Frederic. "Adobe Firefly can now generate more realistic images". *TechCrunch.* October 10, 2023.
techcrunch.com/2023/10/10/adobe-firefly-can-now-generate-more-realistic-images/

[21] John, Daniel. "Artists complain of AI 'copyright infringement' on Adobe Stock". *Creative Bloq*. August 24, 2023.
www.creativebloq.com/news/adobe-copyright-ai

6.3:

It is likely non-copyrightable and public domain works are among the first to be harvested for training generative AI, as they pose fewer copyright or ethical concerns and can legally be used freely. AI models are capable of accurately recreating famous works like the '*Mona Lisa*', for instance. However, public domain works, at present, are less desirable for generative AI- as

most of these works are at least a hundred years old and do not fit modern aesthetics or tastes. This is in part why generative AI companies may heavily rely on web scraping and the use of copyrighted works to make their models marketable, especially in areas like sci-fi and fantasy illustration.

Should copyright protect the works of artists from being fed into generative AI models before falling into the public domain, it will give artists and their heirs the necessary time to benefit from their creations as copyright law intended.

6.4:

From what I have witnessed of AI outputs, there is a degree of retention of the underlying data, like copyrighted works, within AI models. While downloading and processing 'training material' like whole images is necessary in the training process, AI developers often claim that they were done only 'temporarily' are are so 'compressed' in the finalized models that exact copies do not necessarily remain. However, language and image models can repeat specific combinations of text or recreate images when instructed to do so, like reciting passages of popular books or outputting accurate depictions of existing paintings. This means that there must be a form of retention in the models themselves.

Another form of material retention is the collection and publication of datasets. Some datasets may include text, audio, and visual files wholesale (which would violate copyright), but dataset curators like those behind LAION have attempted to skirt around accusations of infringement by publishing 'links' and 'clip previews' instead. [22] These links are akin to links shared on social media- where the copyrighted media is hosted elsewhere, but the links display previews and points to them. Such datasets and dataset hosts may claim their collection of links and image embeddings is not infringement, even though they are providing a convenient list for others to download and utilize the media those links point to for AI training- thus encouraging infringement.

It is also possible training materials may be downloaded to and retained for longer periods of time, if not indefinitely, on private drives or cloud services as a reference to what was employed in AI models, to sell or offer to other developers, to be curated and edited by data workers, or to be used for future training runs. This is most definitely infringement, as it requires copies being made and the hosting, distribution, and use of those copies. However, there may be an incentive for AI developers to discard their dataset files entirely so that there is no record of what was used in the training process. This is why laws requiring data transparency are so necessary in determining copyright infringement and privacy concerns in datasets.

[22] Xiang, Chloe. "A Photographer Tried to Get His Photos Removed from an AI Dataset. He Got an Invoice Instead." *Vice*. April 28, 2023. www.vice.com/en/article/pkapb7/a-photographer-tried-to-get-his-photos-removed-from-an-ai-dataset-he-got-an-invoice-instead

7.1:

My understanding of the training process for AI models (particularly image generation models) is that existing works are downloaded, collected, and labeled with keywords akin to social media hashtags. For instance, the *Mona Lisa* could include tagged descriptors like, 'woman, brunette, long hair, dark eyes, pale skin, Renaissance, by Leonardo da Vinci, oil painting, Italian, portrait, 1500s dress'. Note that these keywords are similar to what AI users input as 'prompts' when generating derivative outputs. The *Mona Lisa*, *Girl with a Pearl Earring*, and even photographs of women's social media profile pictures would be categorized together so

that when a model 'trained' on those works is prompted with 'portrait of a woman', the AI model then 'fuses' (though AI developers call it 'diffusion') 'noisy', pixelated characteristics of the portraits to produce derivative outputs.

Specific images can be reproduced faithfully by AI models, especially if there are fewer images to draw upon of a certain keyword or if, on the opposite end, several copies of a particular image are present and skew a descriptor. [23] Famous paintings and photographs that can be recreated faithfully, according to users on social media platforms like Twitter/X and Reddit, include *Starry Night* by Van Gogh, the *Afghan Girl* photograph by Steve McCurry, and the aforementioned *Mona Lisa* by Leonardo da Vinci.

Additionally, AI models can correlate the 'style' of particular artists through patterns within their downloaded works. For example, a generative AI model could be tasked to produce images of cats 'in the style' of Dr. Seuss by drawing upon its data corresponding to cats and Dr. Seuss illustrations. The AI model in this instance would apply the inked lines, 'Who'-like facial characteristics, and flat colors associated with Seuss onto the form of a cat. However, if a model's dataset did not contain a single image made by or labeled 'Dr. Seuss' it would have nothings to draw upon- and might even output a cat that is dressed like a doctor.

In part or in whole, the images used in 'training' AI are likely retained faithfully within their systems base upon their ability to reproduce images, likenesses, and styles on command.

This use clearly violates the rights of copyright holders, as their copyrighted media must be 'copied' and input into AI systems for analysis and categorization in training. Additionally, the use of artists' works in AI models appropriates uncompensated labor without license or credit to bolster the capabilities of the AI models. The ingestion of copyrighted works for training in AI models is a violation in the mechanical sense (downloading, creating copies), and the spirit of the law (using the works for commercial gain and to the detriment of artists whose works were appropriated by generating competing derivatives).

[23] Xiang, Chloe. "AI Spits Out Exact Copies of Training Images, Real People, Logos, Researchers Find". *Vice.* February 1, 2023. www.vice.com/en/article/m7gznn/ai-spits-out-exact-copies-of-training-images-real-people-logos-researchers-find

7.2:

I am not familiar with the precise digital form 'inferences' take and how they are stored within AI models- but image generators seem to retain potentially pixel-perfect copies of photographs and artwork based on the aforementioned instances where it is possible to have them recreate specific paintings and photographs. Diffusion models use a form of interpolation between multiple images by adding pixelated 'noise'. The assembly of this information in pixels, even if they are 'compressed' versions of the original images, implies the training images are retained and utilized in the generation process.

As for inferences- if a model was, for example, provided only white horses as training data it would result in the generation of horses with white fur by default. Similarly, if 'trained' on stereotypical depictions of people of different ethnicities and nationalities, it will maintain those stereotypes in outputs. [24] This is sometimes referred to as 'bias', but it highlights that training data, including copyrighted works, leave a lasting influence in the 'inferences' models become encoded with.

A team organized by professor Ben Zhao at The University of Chicago has been making defensive tools for artists to alter how their works are used as stylistic inferences in AI with 'Glaze'. More recently, they have made another to shift tagged word associations within AI models that scrape unlicensed artists' works with an upcoming tool called 'Nightshade'. For example, Nightshade would allow an artist to apply a filter to a painting they made of a dog so that when an AI model is 'trained' off of it, the AI would 'see' the image of a cat instead. The researchers demonstrated that even adding as little as 50 images with Nightshade applied would begin to degrade AI models like Stable Diffusion so that it would begin to output hybrids of cats and dogs when prompted to output a dog. If a few hundred images using Nightshade are ingested into AI models, the models would then consistently begin to output cats instead of dogs. [25]

This is convincing evidence that the 'inferences' generative AI make are not a form of knowledge or understanding- but merely a function that draws information from images tagged with the label it is prompted to output.

[24] Harrington, Kayla. "An AI Program Imagined Barbies From Around the World and Wouldn't You Know It, They Turned Out Super Racist". *The Mary Sue.* August 4, 2023. www.themarysue.com/ai-created-racially-stereotyped-barbies-from-around-the-world/

[25] Heikkilä, Melissa. "This new data poisoning tool lets artists fight back against generative AI". *MIT Technology Review.* October 23, 2023. www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/

7.3:

As of yet, it seems impossible for AI models to 'unlearn' information gained from specific pieces of training material. Once assimilated, the creative works and their influence cannot be removed. To remove even one piece of art from the 'training data' may warrant erasure and retraining to ensure that the model does not retain any trace of the work. [26]

This is why it is so important that all media input into generative AI datasets and models be vetted for safety and licensed properly with the express permission of copyright holders and private citizens before training ever commences.

[26] Pastis, Stephen. "A.I.'s un-learning problem: Researchers say it's virtually impossible to make an A.I. model 'forget' the things is learns from private user data." *Fortune.* August 30, 2023. fortune.com/2023/08/30/researchers-impossible-remove-private-user-data-delete-trained-ai-models/

7.4:

Without a full and easily accessible catalogue of creative works used in a dataset, it may be impossible to identify individual works within a model. Instances where there are enough copies of a singular piece that visibly similar images are generated (as mentioned prior with the *Mona Lisa*) is called 'overfitting' by AI developers. 'Overfitting', and the ability of a model to create a deepfake or mimic an artists' 'style', can implicate models in being trained off of copyrighted images.

If the outputs do not closely resemble individual creative works, it does not mean those works were not used in the model to generate outputs.

8:

I believe that using copyrighted material should always require an explicit agreement with the copyright holder prior to dataset compilation and AI training. The fourth factor of fair use is that the result of the use does not compete with the original copyrighted work. For many generative AI applications, the media it produces competes with the media of the underlying data (including copyrighted works), and therefore should be considered infringing if that media was used without creators' consent.

An example of a potentially useful and non-infringing application of AI that I have used myself is Cornell University's machine learning 'Merlin' bird identification app. This app listens to birds you record with your phone, and then identifies the bird based on its calls by comparing them with recordings from the eBird database. This database is constructed from recordings archived by birders at the Macauley Library of Cornell's Lab of Ornithology. While a similar app in this vein might use unlicensed audio of bird calls recorded by nature videographers, the app does not create new audio or video of the birds to compete with those videographers. It would merely be used in bird identification.

However, I would argue that even if apps like Merlin are free and non-commercial, the videographers whose audio made them possible should still be compensated for the work they put into traveling to, finding, and then recording the birds with their equipment. If a nature videographer's bird recordings are often licensed, distributed behind a paywall, or hosted on a platform that pays for impressions and views on the recordings (like Youtube), apps trained off of their recordings would be depriving them of revenue. The presence of the Merlin app has no doubt disincentivized me from seeking out bird calls on Youtube to help in identifying birds, for instance.

So while there are uses for AI that do not directly compete with creators' works and therefore may not infringe in ways defined by fair use- there is still labor in the creation of the media being used as the 'training data' that could afford creators additional sources of revenue. Cornell's Merlin is an excellent example of responsible AI made from their own library, but there may be other bird identification apps that use Youtube videos and other unlicensed sources for their data instead.

8.1:

The rulings of *Google LLC v. Oracle America, Inc* (2021) and *Andy Warhol Foundation for the Visual Arts, Inc v. Lynn Goldsmith, et al* (2022), would support the claim that the use of creative copyrighted works for the purpose of training generative AI is not 'fair use'.

In Google vs Oracle, it was asserted that Google's use of Java SE APIs and code to develop its Android mobile operating system was a means to interface with ubiquitous Java functions and develop compatible apps. In Justice Breyer's analysis of 'the purpose of the use' of Google implementing Java code, he said that the use allowed "programmers familiar with the Java programming language to work with its new Android platform". The Supreme Court sided with Google, as the purpose of the use was as a written function in computer operations to promote interoperability for new mobile devices and interfaces. Justice Thomas and Alito dissented on the decision, however, with Thomas saying that the ruling would result in the potential loss of copyright protections to declaring code. [27]

I do not believe the ruling in Google vs Oracle gives AI developers the right to appropriate creative copyrighted works as Google appropriated Oracle's Java code, however, because the outputs of generative AI fulfill the same purpose as the ingested copyrighted works. Most

importantly, creative works are highly variable forms of human expression and are not instructions, mechanical functions, or concrete facts in the way that code can be interpreted. The Google vs Oracle ruling specifically pertained to declaring code, and not to art and other forms of creative expression.

I agree with the Supreme Court's ruling in favor of Goldsmith in Warhol vs Goldsmith, which involved the licensing and distribution of a magazine cover that copied one of Goldsmith's photographs of the singer, Prince. Andy Warhol, aptly referred to as an 'Appropriation Artist', without license utilized photographs of celebrities, plants, animals, and objects alike to make his iconic pop culture screen prints. For instance, a separate case he settled out of court in 1966 involved the use of Patricia Caulfield's photographs for his *Flowers* series- in spite of the fact that she had refused to license the works to Warhol prior. [28]

Warhol's proclivity to use copyrighted photographs without consent circumvented the licensing market of the photographers whose works he appropriated. Warhol and the AWF deprived photographers like Goldsmith of licensing revenue first in Warhol's unlicensed use of their works, and secondly through the licensing and sale of his derivative works to various clients and collectors. As Warhol's print *Orange Prince,* the impetus for the case, could and had served the same commercial purposes as Goldsmith's original photograph, it was not ruled as 'fair use' by the Supreme Court.

Though outputs of generative AI may outwardly seem aesthetically 'transformative', the 'purpose of character' of the outputs is the same as the copyrighted works (their purpose as usable writing, photography, art, music, voices, etc. as marketable assets or designs for sale or license) ingested and appropriated by AI developers to train, pre-train, and fine-tune their products. Thus, they are not transformative in use and purpose. At all stages of training, generative AI fed with unlicensed copyrighted images infringes via the use of illicit copies, but also due to their intended purpose.

Additionally, I might argue that generative AI models themselves are commercially viable products built off of appropriated copyrighted works- products that serve the same purpose as hirable writers, photographers, artists, musicians, and actors. At their core, such models are artificial 'content creators' 'trained' off of human creators' works and 'prompted' to output what clients ask of them. The Copyright Office may look at the outputs of generative AI as the derivative product in direct competition with infringed copyrighted works- but there may be merit in framing the AI models themselves as derivative products that serve to replace the purpose of artists, creators, and copyright holders.

There are machines throughout history that have displaced human workers- but they were incapable of consuming the fruits of human labor and intellectual property to bolster their effectiveness. Generative AI, even among other AI applications, is unique in that it uses the ingestible identities (face and voice), alongside the intellectual property of human workers and the products of their labor to formulate outputs that mimic humans and their creations. The outputs of generative AI compete with the copyrighted works of creators in markets and licensing, and then compete directly with the creators themselves as commissionable collaborators.

An example of this in the medical field would be the potential for generative AI 'therapists'. Generative AI models are already being marketed as potential replacements for therapists in that they could converse and give advice live to clients. [29] Presumably trained on books, papers, and conversations authored by human therapists and other texts, the outputs of such a generative AI does not just compete with therapists' copyrighted works. In practice,

'therapist chatbots' could be deployed to interact with clients online, effectively replacing the therapists themselves by being trained off of their literary works and recorded sessions.

Simply put, the purpose and character of the use of copyrighted works in training generative AI is to also replace the need to hire creators and copyright holders for the commission of new works or the purchase or license of their existing works. And, even beyond the arts, to replace human beings entirely. Such a purpose also cannot be interpreted as 'fair use'.

[27] Carlisle, Madeleine. "How Google's Big Supreme Court Victory Could Change Software Forever". *TIME.* April 6, 2021.
time.com/5952718/google-oracle-supreme-court/

[28] Bandara, Pesala. "Nature Photographer Who Sued Andy Warhol Over Her Photo of Flowers Dies". *Peta Pixel.* September 11, 2023.
petapixel.com/2023/09/11/nature-photographer-who-sued-andy-warhol-over-her-photo-of-flowers-dies/

[29] Ducharme, Jamie. "Can AI Chatbots Ever Replace Human Therapists?". *TIME.* October 4, 2023.
time.com/6320378/ai-therapy-chatbots/

8.2:

Entities that collect and distribute copyrighted material as 'data' for training AI should not be allowed to skirt around legal consequences for their role in enabling and perpetuating infringement. By collecting and distributing copyrighted material in datasets intended for training AI, they facilitate infringement in downstream uses and applications. Depending upon how the contents of the datasets are stored, the act of copying copyrighted media is also in itself infringing.

8.3:

If there are non-commercial or research datasets and training for AI models that are later repurposed commercially and for profit, they should lose any fair use protections. For-profit AI developers should never fund or invest in non-commercial or research datasets and models, as it calls into question the motives behind the research. As mentioned prior, Stability AI helped in funding the creation of the LAION-5B 'research' dataset, which is one of the image datasets widely used by for-profit and commercial AI developers (including Stability AI itself through Stable Diffusion and other downstream apps like DeviantArt's 'Dreamup'). My own artwork, as mentioned prior, is present in the LAION-5B dataset without my consent.

Non-commercial and research generative models, however, would still violate the rights of creatives and the general public if made widely available due to downstream harms presented by AI users and infringing outputs. Non-commercial and research generative AI could still produce outputs that can be used commercially, and therefore should not be afforded fair use protections if distributed to third parties.

8.4.

AI developers may require millions of examples of creative works to train their models. While it can be argued that if a creative's work is one of many and that their individual influence may be minimal, AI developers do add 'weights' that push specific materials to the fore- often for aesthetic or practical purposes. Regardless of how integral a single piece is to the overall

functionality of a generative AI model, all pieces are still being utilized in such a way that they may collectively compete with their creators. Additionally, while a single piece is one of millions in a dataset, the piece may only be one of tens of pieces a creator makes in their lifetime, and is therefore of great import. The size of a single artist's unwilling contribution to an AI model does not negate the harms posed by the model and its outputs.

8.5:

Under the assumption that the fourth factor of fair use is violated by generative AI (that it is infringing for AI outputs to compete with original copyrighted material it utilized as inputs), the effects of an AI model on market value should be considered on market and class scale. While there are 'image to image' and 'fine tuned' models that exist to specifically edit, modify, or attempt to mimic the style of a copyrighted piece and a single creator's works, those are more egregiously infringement, as they are made with a singular image or artist in the mind of the infringing AI developer or user. These uses of AI are more obviously in violation of the fourth factor.

However, the Office's inquiries into the detrimental effects of AI and the inclusion of copyrighted works in models should encompass a market and class scale because there are relatively unknown creators that may never be specifically targeted by 'image to image' or 'fine tuning' who will still unwillingly have their works utilized by AI. If anything, these artists may be detrimentally impacted more severely by the presence of generative AI because they may be less established in their careers.

Regardless of whether the output of generative AI resembles a particular artist's works, the AI outputs using the sum total of all artists' works compete with the entire class.

9.

Copyright owners should be able to affirmatively consent ('opt-in'), or dissent to the use of their works in AI training before they are utilized. These permissions should not be buried in the Terms of Service of online platforms or similar agreements that obfuscate the use of copyrighted material for training AI in legalese and vague descriptions. Copyright holders should decide for themselves whether or not they want to contribute to or withhold their works from AI developers just as they would a publisher or any other licensee.

'Opt-out' would unfairly require copyright holders, rather than those collecting and utilizing their works without consent, to bear the burden of discovering and then reporting the misuse of their works. With AI developers currently building hundreds of models based upon datasets that do not respect consent by utilizing web scraping to acquire their data; it is possible those advocating for 'opt-out' do so with the intention of creating, launching, and profiting from generative AI products before copyright holders can protest.

AI corporations and developers cannot be trusted to act on 'opt-out' requests, especially if honoring them would degrade the quality of their AI products.

As mentioned prior, there may be no means to make a model 'unlearn' a specific piece of data, which means 'opt-out' is unsound as a means for artists to prevent their work from being used in AI applications. The responsibility of data curation and seeking permissions should fall squarely on dataset makers and AI developers as they are the ones with immediate access to the datasets and are using the works as the building blocks for their products.

9.1:

Consent should be required for all uses of copyrighted works in training AI models, not just in commercial models. This is because there are several 'open source' and 'free' generative AI models that can produce outputs that are still capable of competing with original copyrighted works. For instance, if an image generator is 'free', yet its users take outputs to create commercial products or even non-commercial images that they otherwise would have had to pay artists to produce- they are still profiting from the uncompensated labor of copyright holders. Non-commercial generative models would still detrimentally impact the market for creators by allowing users to unfairly profit off of their works.

9.2:

From my personal experience issuing DMCAs for over a thousand instances of infringement prior to the development of generative AI- I can attest that adopting 'opt-out' measures will only add to the stresses and burdens independent creators have to endure to assert their copyrights. The Copyright Office and US government should mandate 'opt-in' only instead.

If, to disastrous effect, 'opt-out' is decided upon, every AI model in existence would have to have a publicly searchable dataset documenting every piece of work incorporated into their training data. There would have to be third-party auditors to ensure that every piece of data is accounted for, so that AI developers does not discard or hide data. Copyright holders would then have to be aware of every single AI model, find their dataset catalogs, and then have to use text inputs or image detection tools to scan the databases and issue reports. Many copyright holders and their heirs may not even be US citizens, understand English, or have a means to become aware of the datasets' existence.

Even if such a system saw mass adoption and 'opt-outs' are successfully registered, the only way to ensure the works are removed from the infringing AI models would be to destroy the models entirely- each and every time a new infringement is reported.

'Opt-Out' is an inefficient, ineffective, and cumbersome process. I have attempted to 'opt-out' hundreds of images of my art and thumbnails from a site called 'HaveIBeenTrained' by Spawning, which is supposed to log 'opt-outs' for the LAION-5B dataset and Stable Diffusion models. From my personal experience, such sites are poorly built, slow to load, do not accurately make a record of 'opt-outs', and may not find data even with basic tags, titles, site information, or author names as search terms. Some developers, like OpenAI, request that artists even upload their works to cross-reference for 'opt-outs'. Artists like myself would never willingly provide our work if there was the possibility they may then harvest the images for training regardless.

And even when 'opted-out', images may still available to other AI developers and individuals to view and download through the logged and distributed datasets. The potential lack of 'opt-out' compliance on the part of AI developers, and their unwillingness to destroy models already in circulation, also defeats the purpose of the whole endeavor.

Metadata, while useful, is often stripped and nullified by social media platforms upon being uploaded and posted. Each subsequent copy, in the case of images and video, would then be shared without any metadata or even incorrect or edited metadata. Additionally, not every digital program has an easy means to add metadata to creative works, and creators may not be familiar with the process. The presence or lack of metadata should not be used as an indicator as to the viability of using a work as training material.

Copyrighted works may have watermarks or notices written on them, but they are often

flagrantly ignored or even cropped out by infringers. Infringing uploads of my own paintings often had my watermarks and copyright information removed. There are even AI models made with the express purpose of removing watermarks being released in recent months. Even if sites or files are coded and labeled with copyright notices or robots.txt commands, there is no guarantee they will be respected by scrapers or AI developers.

Unless there are penalties for ignoring crawler tags and erasing metadata and notices, AI developers will likely continue to scrape works that have been stripped of their copyright information. That, or they will utilize datasets assembled by third parties and therefore can deny knowledge or responsibility.

Thus, 'opt-in', with due diligence to ensure that the copyright holder is in fact the one 'opting-in' their intellectual property, is the only way to ensure that AI does not unfairly benefit from unlicensed copyrighted works.

9.3:

As mentioned in 9.2, there are too many points of failure in using 'opt-out', notices, or metadata to ensure that copyrighted works are not harvested, ingested, and exploited by generative AI models and their developers.

I am not aware of studies to see how few datapoints a generative AI model can draw from before its outputs visibly copy underlying data. Most of the models on the market used without license millions, even billions of works as data. If millions are in fact required, it may be difficult for some AI developers to attain consent from every copyright holder whose work they wish to use in their models.

But to put this issue of data access into perspective- chefs are not forced to cook for free, even though food is a human necessity, nor the farmer required to give their produce to the chef without payment. A gourmet chef may produce higher quality meals than other chefs due to a combination of their extensive training, high quality equipment, and the rarity of the ingredients they purchase. A chef's competitors are not guaranteed access to the same high quality training, equipment, and ingredients unless they can pay for them as well.

Artists and other creators also work within these confines- with training, supplies, mediums, and tools that are within their financial means to acquire. An artist could not justify stealing supplies, even if the quality of their work could be improved with them.

As such, copyright holders, including artists, should not be forced to give their works to AI developers so as to subsidize the quality of their generative models for free. It is unfair to require that artists and creators, who frequently struggle economically, give the fruits of their labor to AI developers (large or small) to create competing products and services. Like any other business, AI developers must pay for what they wish to acquire for their products.

9.4:

If an objection to the use of their work in training AI is not honored, copyright owners should be able to receive statutory infringement fees, sue for royalties for the past and continued use of their work in existing models, and even request the removal or destruction of the AI models with the aid of a governing body like the FTC (via 'algorithmic disgorgement').

Unfortunately, the current legal system is slow to dispense justice for copyright holders, and it is likely AI companies will have accrued enough revenue and investment to effectively block litigation against them. Thus, a large subset of copyright holders will not be capable of fighting

for their rights in court. Years may pass before current cases against AI companies are litigated and set precedent, which will unfortunately allow millions of copyrighted works to be used time and again in the development of AI models without license in the interim. [30] [31] We cannot rely on the courts to shape policy swiftly enough to counteract the present harms generative AI poses to copyright holders and the public.

There will need to be new remedies. It may be underway, but I would suggest that the Copyright Office make the official registration of human made copyrighted works and the creation of modernized online 'Copyright Holder' accounts through the Office swifter and easier. I myself have registered multiple pieces of my artwork with the Office, but artists and other creators often publish several pieces per month on social media. Some creators post new works daily, even as partially finished 'works in progress', and cannot wait for registrations to be processed before publication. If they are not 'unpublished' and registered collectively in batches, copyright holders are expected to register individual 'published' works at $55 a piece.

If I had registered every painting and drawing that I created in a single year, it would be thousands of dollars to register them all. That is prohibitively expensive for copyright holders, many among them independent artists and teenagers first publishing their works online. This pricing scheme often forces many creatives to do without registration entirely, which harms their ability to litigate when their works are infringed upon.

Artists, thus, should be able to register works in batches, even when they have already 'published' the works online to market themselves. I may also suggest removing the 'timely manner' limitation that requires pieces be registered within 3 months of creation/publication in order to litigate. The ability to litigate against an infringer should not be predicated on whether or not a work is registered within a particular timeframe, as the copyright still belongs to the original author during their lifetime and for a period thereafter.

If the US Copyright Office is successful in modernizing their site and registration platform while making the process intuitive and more affordable to artists, it will do much to encourage creators of all disciplines to register their works and protect their copyright.

It may be an outlandish suggestion, but it may also be possible for AI model developers to 'register' the contents of their datasets as a form of collective work in a database for datasets that can be cross-referenced with the Copyright Office's records. If the system detects a copyright holder's works registered with the Office in one of these datasets, the copyright holder and Office could be notified. The Office, through a body similar to the claims court established by the Case Act (but with no option for the infringing AI developer to avoid proceedings) could dispense judgement, award damages, and request the deletion of the model to the FTC via 'algorithmic disgorgement'.

This could provide a means for the Copyright Office and the public to be able to detect and arbitrate infringement perpetrated by AI developers at least within the United States on behalf of registered US copyright holders. This would also give creators a means to defend their rights without having to begin long and costly Federal proceedings.

[30] "From ChatGPT to Deepfake Apps: A Running List of AI Lawsuits". *The Fashion Law*. October 19, 2023.
www.thefashionlaw.com/from-chatgpt-to-deepfake-creating-apps-a-running-list-of-key-ai-lawsuits/

[31] Sheng, Ellen. "In generative AI legal Wild West, the courtroom battles are just getting started". *CNBC.* April 3, 2023. www.cnbc.com/2023/04/03/in-generative-ai-legal-wild-west-lawsuits-are-just-getting-started.html

9.5:

Yes, creators should reserve the right to deny AI training on their 'work for hire' pieces that technically belong to another entity. Many artists in the film and games industry participate in 'work for hire' employment, which gives the studios that hired them exclusive rights to the works thereafter. Such artists are the most at risk for displacement by generative AI, as they may have produced enough work for clients that, should the client license their work to an AI model, they could feasibly never hire the artist again but still utilize their 'style' and the quality of their work.

Most importantly, a human creator should have every right to refuse AI training on their work because the 'style' and expression in the work belongs to the artist. The expression is not the client's, and therefore not theirs to give or license to AI developers.

Clients who own 'works for hire' should keep information on the creator of the work, and must seek the consent of the creator before licensing the work for use in generative AI.

10:

Licenses should be obtained emphatically, plainly, and with proper negotiations.

Social media platforms and other services (particularly established ones copyright holders and individuals have been using for years, like Meta's products, Microsoft, Apple, Twitter/X, and the like), should not use their Terms of Service to gain access to all data on their platforms. These platforms should also not demand that deletion be the only means to remove or protect users' works from exploitation. While some sites offered 'opt-out' designations for accounts and even individual pieces or media, the default for users should always be 'opted-out' so that they manually have to 'opt in'. The default for works posted to DeviantArt, for instance, had been 'opt in' before outcry by artists who had to manually opt each piece in their portfolios 'out' convinced the platform to allow for account-wide and default 'opt out' settings. [32]

Sites should not force their existing users to supply data for AI models, especially if the sites have only recently started to invest or develop AI and it was not the primary product when its users initially signed up for their services.

[32] Foley, Joseph. "DeviantArt breaks silence over fierce backlash for its new AI art tool". *Creative Bloq*. November 15, 2022. www.creativebloq.com/news/deviantart-dreamup-ai

10.1:

'Feasible', when addressing licensing, is less about capability and more about ease and expense- and whether the penalties for using unlicensed work appropriately harm infringers. Voluntary licensing will no doubt be difficult when millions of pieces of copyrighted works are desired in the development of commercially competitive AI models.

However, with the value of AI businesses being estimated in the billions, voluntary licensing will be the only way individual copyright holders that are not part of a collective can extract that

value from the entities that seek to repurpose their work and potentially displace them in the market.

10.2:

The Music Modernization Act is an example of collective licensing that was once deemed infeasible. Prior to this legislation, sites and streaming services were distributing illegally downloaded copies of musicians' works without license or compensation. The music industry could have collapsed were it not for the Act demanding renumeration and licensing. Music, like art and writing, is already being used by generative AI models, and should be protected.

Sites resembling music streaming services have cropped up in recent years with licensing schemes for art, photography, writing, and more. Stock photography sites like Adobe Stock, Shutterstock, and Getty Images are examples of entities licensing imagery on a collective basis. They have even partnered with AI developers to make models based upon their stock offerings- though, as mentioned prior in my answer to 6.2, it is contested by contributors that consent was sought before training commenced.

It is completely feasible that there are platforms that will be able to provide properly licensed material to AI developers. An equivalent of streaming and stock sites could be established and voluntarily joined by creators who want to license their works for the purpose of AI training and receive a share of models' profits. However, additional precautions and checks should be implemented in the creation of collective and stock licensing schemes, as bad actors and infringers can always upload media they do not own.

It is also possible that collective licensing schemes, like mentioned prior with The Music Modernization Act and stock site AI contributor payouts (like those offered by Adobe, Getty, and Shutterstock), will not be sufficient to support creatives compared to the licensing they offered individually for their works prior to generative AI. After a survey of over 50 Shutterstock contributors, for instance, it was found that on average they were being paid 0.0078 cents per work. [33]

Artists and other copyright holders should never be forced into collective licensing schemes, lest the compensation determined by the entities using their works be unfair and insufficient.

I do not believe the United States should adjust anti-trust laws to allow for monolithic entities made for collective licensing. This would discourage competition between licensing platforms, and likely result in low payouts for licensed copyrighted works. If multiple platforms compete, it will hopefully increase revenue for artists who choose to license works through them.

[33] Wiggers, Kyle. "How much can artists make from generative AI? Vendors won't say". *TechCrunch.* September 30, 2023. techcrunch.com/2023/09/30/how-much-can-artists-make-from-generative-ai-vendors-wont-say/

10.3:

No. It is not fair to compulsively force creatives to license their works, which are so personal and unique to them and required years of practice and labor, to entities that will repurpose their creative expression to develop products that will later compete with them.

Art, writing, music, and other creative products, while integral to society, are not required to live like access to food, clothing, shelter, and medical care. We do not compulsively require that US

citizens redistribute their personal property (their food, their clothing, their homes etc.) to others. The same should be true for creators and their intellectual property. It would be hypocritical if artists and other creators, who are often small business owners themselves, were forced to provide their labor and products to AI companies while other businesses do not.

For example, the United States would not require Coca Cola forfeit its recipes to other soda companies, Apple to give its operating system up for use in non-Apple phones and computers, or force taxi services to give their vehicles to Uber drivers. I can conceive of a thousand examples in which a company would have the right to withhold their property or innovations from competing businesses and services.

Compulsory licenses, particularly for the creations of often abused and exploited artists and middle class workers, would be cruel and unjust when the primary beneficiaries are powerful and monied AI companies.

I believe no mandatory fee under a compulsive license would be able to compensate artists for the violation of having their creative works used against their will by their competitors, nor the destruction of jobs and opportunities in creative markets and industries.

10.4:

No. Extended collective licensing schemes should not be allowed, as they would bypass the consent of individual artists and creators in a copyright class. Unrepresented members should never be forced to license their work for AI development under extended schemes.

10.5:

Licensing should not be varied based on the type of creative work, as having permissive licensing for some forms of artistic media but restrictive licensing for others would disincentivize creators from crafting media with more permissive licenses. Every creator should be afforded the same protections for their works, whether written, drawn, or sung.

11:

Licensing will be difficult, but necessary for developing lawful, fair, and ethical AI products. The volume of works required for training, finding the copyright holders associated with text, audio, and images online, and then corresponding with those rightsholders all take time and due diligence. This is why, as mentioned prior in regards to collective licensing in my answer to 10.2, the process can be expedited with centralized platforms made with the intent of offering licenses to AI models.

Data curators should always seek licenses before adding media and data to datasets, as these datasets becomes the training material that is used by AI developers and the companies that employ AI models. AI developers and companies, as the primary beneficiaries of the datasets, may need to be involved in funding datasets so that licensing can take place. If AI developers or companies seek more specific 'data' for their models, they may need to facilitate the licensing of copyrighted works and media that meet their requirements. Thus, at all levels, each member of the chain needs to be responsible for licensing and ensure that the datasets they are using are sourced with properly licensed works.

If infringement takes place at any step in the chain, the whole system would be infringing. If the final models also facilitate infringement by allowing for the additional input of unlicensed material via 'fine tuning' and 'image to image' functions, the developers and companies that deploy them should also be held liable for downstream abuse.

Non-commercial and research datasets should be licensed as well, especially when applied to generative AI systems. They must also ensure that these datasets are never released for commercial purposes, to business entities, or to the public due to downstream harms.

12:

As of yet I have not heard of means to determine to what degree a single work contributes to a generative AI output. It is possible if a model generates an image of a 'dragon', one would assume all images of dragons from the dataset could have contributed to the image more so than images of flowers. However, it could have also used photographs of lizards and other dragon-adjacent images as well. With more tags used, there is no knowing how many images could contribute to a single output. It is also suspect that AI models would be capable of performing in the reverse and could identify what images were used in an output.

Visual similarities between an output and images in the dataset are subjective, and thus cannot be used as an approximate measure.

13:

While a licensing requirement would potentially slow the development of generative AI and make it a more costly venture- it can be argued that the resulting AI models will have their underlying data properly vetted, and therefore be safer for public consumption.

If anything, licensing requirements would give writers, musicians, artists, photographers, and other creatives additional sources of revenue and a share of the multi-billion dollar evaluations of AI companies that are currently profiting off of their unlicensed work. AI companies could benefit from exclusive licenses with specific creators, bringing demand to their model over others'.
Licensing should not be perceived as a barrier to development, but as a beneficial exchange if implemented properly.

14:

Much of this inquiry has rightly focused on copyright and ensuring that copyright holders are not exploited by generative AI which repurposes and competes with their work. However, even should fully licensed models be mandated by the law, there will be entities outside of the United States and 'open source' models being widely downloaded and distributed today that already utilize unlicensed copyrighted material. These models currently harm US copyright holders and citizens, and laws may need to be put in place to ban their use, penalize developers and users, and sanction businesses beyond our borders that abuse US data.

There is also the grim reality that fully licensed models in the near future, and models centuries from now using public domain works from the present day, will still irreparably harm creatives and other workers and their ability to make a living from their craft. It is not unfounded to envision a future where humans no longer need to learn how to write, to play instruments, to draw, to act, to sing. And those creatives who choose to learn the arts will find themselves at odds with AI models and corporations that have centuries of creative works at their disposal and outputs that can be produced in seconds.

Generative AI, even if only 'trained' on licensed and public domain works, is still poised to threaten skilled middle-class workers in computer-utilizing jobs, communication, and creative occupations as employers and consumers use it to cut costs. It is not difficult to imagine that

the widening gap between the wealthy and working poor in our country will grow yet wider as workers are replaced with machines.

The United States may need to consider limiting AI usage in certain fields and hold companies accountable for using AI to displace their human staff. And, to ensure the continuation of human artistry, the US and charitable organizations could support and develop grants, scholarships, unions, platforms, and events that protect, promote, and cater to human-made art and performances.

15:

Yes. So that copyright holders can determine if their works were used to train AI, AI developers should maintain detailed records of every work, file, and piece of 'data' used to train their models. Dataset creators, as the collectors of the data ultimately used to train AI models, must also keep their datasets properly recorded.

Transparency of the underlying data will be necessary in determining whether copyrighted, illegal, or private media and information was used to train AI models. As such, it may be necessary to require these datasets be registered with a body like the Copyright Office so that they are all accessible in one place for government auditors and the public to scrutinize.

16:

Ideally, a mandatory notification and correspondence would be given prior to the collection and use of the copyrighted work so that licensing agreements and permissions can be finalized before the work is added to datasets and 'ingested' or 'trained' into AI models. Such a notification would essentially be a licensing proposal. All copyrighted 'data' should be licensed and used with the consent of their owner.

As there are models currently on the market using unlicensed copyrighted works, I believe notifications should be issued as swiftly as possible so that copyright holders can request the destruction of the models and datasets employed by those models. Ultimately, these existing models that exploit the works of artists should be removed from circulation immediately, then disgorged after notifications are issued and statutory damages and any profit the model may have accrued is distributed to impacted copyright holders.

In short, existing models should remove their models from circulation, notify copyright holders, compensate copyright holders, and then destroy their models. Developers ready to train new models in the future should notify copyright holders before the works are collected and trained upon so that proper licensing agreements can be negotiated.

17:

Outside of copyright law, numerous US laws require disclosures of the materials used to make products. 'The Nutrition Labeling and Education Act' overseen by the FDA require ingredients and nutrition information to be disclosed on food labels. The FTC also enforces Textile and Wool Acts that require fabrics and clothing to be labeled with their fiber contents, often as a percentage, as well as their country of origin.

As generative AI models are fed images, texts, and recordings, it stands to reason that they should disclose materials used in training in a similar manner. There are materials in food and building, for instance, that are illegal or banned substances (like lead and asbestos) that endanger consumers. Training data used in AI development similarly uses illegal forms of media, like private records, non-consensual pornography, and CSAM. [34]

The FTC will likely need to be equipped to oversee the contents of AI datasets so as to protect victims and consumers harmed by AI outputs and unfair trade and business practices. Divulging the underlying data, thus, will be necessary for the enforcement of Copyright and for consumer protections.

The proposed 'American Data Privacy and Protection Act' would demand transparency from entities that partake in data collection (Section 202). It also calls for individuals having the right to access and delete their data (Section 203), withdraw consent to collect their data (Section 204), and maintain the privacy of minors (Section 205). AI models with unlicensed and scraped sources of data used in training may violate the privacy rights of thousands, if not millions, of Americans. Unfortunately, many state and federal privacy bills have not passed in part because of lobbying efforts by Big Tech companies. [35] These very companies, including Google, Microsoft, and Meta, are funding and developing the largest generative AI models now available on the market.

Data transparency must be mandated to not only protect copyright holders from exploitation, but consumers from having their private data appropriated or being harmed by AI outputs.

[34] Tidy, Joe. "Paedophiles using AI to turn singers and film stars into kids". *BBC.* October 24, 2023.
www.bbc.co.uk/news/technology-67172231

[35] Claburn, Thomas. "Big Tech loves talking up privacy- while trying to kill privacy legislation". *The Register.* May 27, 2022.
www.theregister.com/2022/05/27/big_tech_privacy/

Infringement:

22:

Yes. Generative AI outputs implicate the exclusive rights of existing works because they are, by their very nature, unauthorized reproductions and derivatives. This is true in all circumstances, even when the outputs do not bear specific likeness to individual works in the training data, because the AI would be incapable of producing outputs without utilizing the copyrighted works.

As copyrighted works were copied, labeled, ingested, and incorporated into AI models to make competing derivative outputs, AI models and their developers violate the exclusive rights of those whose works were appropriated to train their AI products.

23:

A similarity test is not adequate to address infringement claims of AI outputs. While AI outputs can resemble their underlying data, especially when prompted to do so or given a specific example to alter, the outputs of generative AI can utilize numerous works.

As mentioned in my observations at the start of my letter- dozens, if not hundreds of AI generated images have been uploaded to the web in recent months of the mythological creature I illustrated. It is possible they all used an AI model that trained off of my work without license to benefit AI developers/users and produce competing derivatives. Presumably, when the AI generators are told to make an image of the creature, the generator then accesses

inferences and data taken from my work and every other illustration of that creature it was fed to generate the new image. While the AI generated creatures will bear only a passing resemblance to any one particular example from the training material, the images' quality is dependent on the quality of the images they were fed.

There are, however, more clear cut forms of infringement. 'image to image' allowing AI users to have an AI model modify or copy a specific piece uploaded by the user. Additionally, AI users can 'prompt' using the titles of specific works and the names of specific artists, showing a willful desire to mimic an existing work or an artist's 'style'. In these instances, knowing what additional data was added to the base model and the textual 'prompts' could be used as evidence of infringement.

If a new standard is established, it could be the assessment of the AI models' or their users' access to the infringed works. If the infringed work is present in the AI model's training dataset, the AI therefore had access to the work. If the infringed work bears labels in the dataset that were called upon by the AI to generate outputs via 'prompts' (including titles and the artist's name), the AI likely utilized the infringed image in the process. If a downstream AI user uploads the infringed work to the model, the AI model was thus given access to the infringed work.

This is why AI outputs should always bear labels and metadata to catalog its model, user, 'prompt', and possible reference images. AI models and services may also need to keep records of every image they generate and the output images' associated prompts and inputs to ensure that infringement claims can be addressed.

24:

If AI developers do not maintain records of their training data, it may be difficult for copyright owners to prove models utilized their works without obvious similarities. 'Access' may be evident if the artist displayed their work on a public site that was scraped by dataset makers and later used by the AI developers. A list of prompts recorded by the model or embedded into the AI generated media may also shed light onto potential infringement (like specifying artists names or work titles). Or, if the AI model is 'open source' and accepts user inputs, it is always possible an artist's works could be input after the model's release.

I do not believe civil discovery rules will be sufficient, as AI developers have every incentive to prevent copyright holders and AI competitors from discovering the contents of their training data. This is why it should be mandatory that AI models and datasets have exhaustive lists of all 'data' used, and they should be readily available and confirmed by third parties to ensure every piece of data is properly recorded.

25:

If AI generated media is found to infringe on a copyrighted work, liability will be dependent on who facilitated and profited from the infringement.

In 'closed' AI models provided by companies and services through platforms they control, the dataset curators are most certainly responsible for infringement if they include unlicensed copyrighted works in their dataset. Similarly, AI developers who utilize datasets for training their model should be liable if they knowingly used infringing data or did not properly vet the dataset prior to training. The distributors and hosts of AI models should similarly be held liable if they knowingly use an infringing model or did not properly vet the model and its dataset for infringement. In this situation, end users may not be liable if they were assured the model was not infringing. Users should, however, always be liable if they gave instructions to the model via prompting with the intent to infringe (like specifying an artist or specific piece of intellectual

property). End users should also be liable if they knowingly use infringing models to profit from their outputs.

In editable and 'open' AI models that allow for additional data to be added by downstream AI developers and users, all parties could be held liable for downstream infringement. Dataset curators may not be liable if their initial dataset is completely licensed, but model developers and distributors of editable and 'open source' models should be liable for downstream infringement. This is because 'open' models allow for the addition of unlicensed copyrighted media, which facilitates infringement. Users who specifically 'prompt' the model with an artist/ IP and/or add unlicensed data to the model would be liable for infringement because of their intent to use the AI model to take advantage of copyrighted works.

25.1:

Yes, 'open source' models pose immense challenges with respect to the potential for infringing outputs, as these models are not controlled by a centralized entity that can be held accountable and be properly audited prior to the release of their products.

'Open source' models are infinitely distributable and editable and can be fed additional training material- making it more likely that they will be used to appropriate copyrighted media without consent or license. Early 'open source' models like Stable Diffusion utilize datasets rife with copyrighted media (including my own artwork and photography present in the LAION-5B dataset), and serve as the basis for 'forked' off-branch models utilized in potentially dozens if not hundreds of customized apps.

AI users often modify 'open source' models with the express purpose of mimicking artists, for instance, by feeding them the portfolios of their target. These models (often referred to as LoRAs- 'localized representation adjustment') should most certainly be considered infringing due to users willfully and intentionally feeding the models specific works they want to adjust or emulate.

The harms these 'open source' models pose by using unlicensed copyrighted works cannot be undone, as these models can be distributed freely. Due to the editable nature of 'open source' generative AI models and their ability to be used to harm copyright holders, I believe they should be barred from release and removed from the internet, especially if they already utilized copyrighted works without consent.

26:

If generative AI systems are trained on media that had identifiable copyright information, I believe the outputs would be in violation of 17 USC 1202(b). This code prohibits the removal of copyright information from a work, incurring statutory penalties for infringement.

When an AI model utilizes a work in training and generation, it downscales and distorts the information from the originating images via the addition of pixelated 'noise'. This results in the whole or partial removal and erasure of the copyright information. AI developers and users may be incentivized to intentionally remove watermarks from copyrighted works prior to training to have 'clean' data. This would be a more deliberate violation of 17 USC 1202(b) as well.

Many generative AI models output images with the inferences of watermarks, signatures, and other copyright information. In fact, in '*Getty Images (US), Inc vs Stability AI, Inc*' Getty argues that the presence of their watermark, whether whole or in part, on Stable Diffusion's outputs

demonstrates that Stability not only used Getty's images in AI training, but that it was both removing and adding Getty's marks and information to outputs. [36]

[36] Setty, Riddhi. "Getty Images Sues Stability AI Over Art Generator IP Violations". *Bloomberg Law.* February 6, 2023.
news.bloomberglaw.com/ip-law/getty-images-sues-stability-ai-over-art-generator-ip-violations

Labeling and Identification:

28:

Yes, mandatory labeling of AI generated media should be implemented and enforced.

These labels must be prominent (visible watermarks, audio remarks) and additional notices should be invisible but machine readable (metadata, invisible watermarks, inaudible sounds). Labels could include the date of the output, the models used, and the AI user who prompted the output. Invisible metadata and marks could also include the prompts used to produce the output, so that it can be revealed if the model was instructed to infringe in any way.

Labels such as these will create accountability to models and users for illegal or infringing outputs, and inform consumers as to whether what they are seeing or hearing is AI generated. This will help mitigate the harms posed by deepfakes and give consumers who want to patronize human-made media and products a means to avoid AI generated content.

As generative AI models output generated media, the models themselves could automatically apply labels and metadata. Model makers and AI companies should be required to adopt standards set by the government in labeling should they be allowed to operate in the United States and serve US citizens.
28.1:

AI developers must be responsible for building permanent labels into their systems so that every piece of media output by their models are identifiable.

Additionally, end users like websites, businesses, news outlets, marketplaces, influencers and creators who use generative AI should be required to divulge any AI use and whether their offerings and products, including those by third parties, have been generated with AI. This will help creators and consumers discern between human made and AI generated material, and decide for themselves what businesses to patronize. Some artists and consumers do not wish to purchase or interact with any AI generated products or media, and should be able to, at a glance, identify sites and products that use AI generation in any capacity. This will hopefully encourage some sites and marketplaces to commit themselves to provide wholly human made and authored media and products.

28.2:

There are no barriers to labeling in imagery and footage, as AI models are already capable of outputting images. Adding additional layers of labels should not be technically challenging. Some AI services already implement visible and invisible watermarks and metadata into their models on all outputs like with Adobe credentials and Dall-E marks. However, visible labels should always be enforced to ensure that consumers are properly informed.
Labels on AI generated text and audio (like music) may be more difficult to implement, but are possible as well.

28.3:

AI models that do not label their outputs should be removed from the US market until they comply with labeling mandates. A consequence for failing to label AI generated media should first be the requirement that all copies of the unlabelled media it output be removed from distribution. If the media is ever to be re-uploaded, it must be with proper labels.

Failing to add labels or participating in the act of removing labels should incur statutory financial penalties similar to when notices are edited off of copyrighted works.

Other financial and even criminal consequences for failing to label AI generated material should be determined based on these factors-

-How widespread was the unlabelled media and how many views and interactions did it garner?

-What was the nature of the unlabelled media? Was it illegal (like CSAM)?

-Did the unlabelled media cause confusion, harm, or was used to deceive audiences and consumers?

-How much profit was accrued from the unlabelled media?

-Did the unlabelled media mimic the works of a creator or the likeness of an individual?

29:

Multiple tools have been made in an attempt to identify AI generated material, but they can be inaccurate and are thus abandoned- even by developers like OpenAI. [37]

There are artists who have shared on social media that they put their own handmade works into AI detectors and received false positives that their own works were made with AI. This may be in part because their handmade works resemble the pieces ingested by AI during the training process. As generative AI was made with the intention of outputting convincing images based on human works, it is only natural to assume that models outputs will prove challenging for detectors.

Also, it may be more difficult to identify AI generated text and music compared to AI generated images and video because there are defined methodologies and rules to both that are easier to mimic and recombine.

As such, the best detection tools will be those that can read the invisible watermarks and metadata that will hopefully be mandated as labeling requirements for all AI outputs.

[37] Murphy Kelly, Samantha. "ChatGPT creator pulls AI detection tool due to 'low rate of accuracy'". *CNN.* July 25, 2023.
edition.cnn.com/2023/07/25/tech/openai-ai-detection-tool/

Additional Questions about Copyright Issues:

30:

I believe several rights are infringed upon when AI generated material features the name, likeness, or mimics the voice of individuals. Depending upon the nature of the material being made, it could be libelous or defamatory as generative AI can output convincing facsimiles that can be interpreted as truth. AI can also be used to create false endorsements or allow third parties to benefit commercially from the image of an individual without their consent. It violates moral and publicity rights, and causes harm especially to those who depend upon the exclusivity of their image or voice for their career. AI can generate, in essence, counterfeits of actors, models, voice actors, singers, and other professionals whose voices and likeness are the 'product' they use to make a living.

If these rights are not legally defined as of yet, they should be as swiftly as possible so as to mitigate harm.

31:

Congress should enact a Federal Right to Publicity that would entirely prohibit the use of AI to create visual or audial deepfakes without the explicit consent of the featured individuals for each and every application. While there is the concern bans could halt forms of speech like parody and satire- both have existed for hundreds of years with impersonators and illustrated depictions of public figures. Unlike impersonators and illustrations, which are obviously satirical, AI deepfakes can be indistinguishable from legitimate audio and video and difficult to disprove. AI deepfakes have already been used to create non consensual pornography, facilitate phone scams, and create convincing video of politicians and world leaders issuing deceptive statements. [38] The public needs to be protected from every potential use of their voice and image.

Some states have already begun implementing or considering laws criminalizing the making and distribution of deepfake pornography (CA and NY among them). While deepfake porn is the most common form of deepfake, (a purported 96% of AI generated images and video are pornographic deepfakes, typically of women [39]) deepfakes in any form can detrimentally impact US citizens. Federal Laws should be preemptive and establish a floor for States to abide by and build upon.

It is my hope, however, that deepfakes are banned entirely unless the individual whose voice, body, or face is being repurposed opts-in emphatically for each and every use.

An example would be if an actor allows their face and voice to be cloned with full license and compensation for films that need to 'de-age' them, edit their likeness, or keep a character they portray consistent in future franchised media. Even figures that have been dead for decades should not be mimicked or altered by deepfakes, as it can damage their legacy, taint information streams, and threaten historical record keeping. It is deeply unsettling that laws are being considered that would allow a person's voice and visage to become 'public domain' 70 years after death akin to copyrights. We should not condone puppeteering and disrespecting the dead with generative AI.

Under no circumstance should US citizens have to live with the fear that their photos and recordings will be used to create pornography, scams, disinformation, or profit for unscrupulous businesses- before and after their death.

[38] Steinbuch, Yaron. "Traumatized Ariz. mom recalls sick AI kidnapping scam in gripping testimony to Congress". *New York Post.* June 14, 2023.) nypost.com/2023/06/14/ariz-mom-recalls-sick-ai-scam-in-gripping-testimony-to-congress/

[39] Tenbarge, Kat. "Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy." *NBC News.* March 27, 2023. www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071

32:

Yes, protections are necessary to ensure that generative AI cannot be used as a means to create outputs in the 'style' of human creators. 'Style' is a product of numerous factors unique to every artist- and is determined by not only medium and theme but the imagination, personal preferences, and practiced capabilities of the creator. For artists, their 'style' is often what distinguishes them from their peers, and what draws consumers to specifically purchase their works. While 'styles' cannot be copyrighted, a deliberate effort to copy the hallmark aesthetics of specific artists, especially with AI (presumably by letting the AI ingest the works of the artists to 'train' on their work and produce derivatives) should not be tolerated.

Should AI be used specifically to mimic the works of artists, it would only serve to create, in essence, highly accessible and infinitely producible counterfeits and forgeries. AI would allow for individuals and corporations to generate media in seconds that they otherwise would have commissioned, licensed, or purchased directly from creators. Additionally, a proliferation of AI media 'in the style of' an artist would saturate the market and cause confusion among consumers. This could be damaging to the credibility, reputation, and marketability of creators.

Artists have already observed their names and 'styles' being co-opted by AI generated media in online search engines, with even paintings by the greatest artists of the past few centuries being outranked in search results by AI facsimiles. This occurs especially when the artists' names are used in the 'prompts' to generate AI images. Greg Rutkowski, a fantasy artist from Poland known for his work with US game maker Wizards of the Coast, was among the first artists to be targeted by AI users. In 2022 alone tens of thousands of images were generated with his name in the span of a few months via the Midjourney and Stable Diffusion models. He expressed his dismay in that his own artwork was being displayed in search engines next to AI generated counterfeits using his name- undoubtedly causing confusion and unfairly associating him with images that he did not create. [40]

Another artist, Sam Yang, discovered that AI users had made specific 'fine-tuned' models mimicking his colorful portraits of stylized women. When confronted, the AI users continued to generate images in Sam's 'style' and even used 'image to image' AI to alter his existing works. [7] I have witnessed my fellow illustrators recount stories on social media of having their works directly fed into AI models by end users seeking to acquire images in their unique 'style'. Some of these AI users even use AI models to create lewd or disturbing outputs maliciously to attack artists who attempt to defend their work.

AI users similarly prey on recently deceased artists' styles. An example of this occurred after the untimely death of renowned Korean illustrator and manhwa (Korean graphic novel/comic book) artist Kim Jung Gi in October of 2022. He was best known for bustling scenes of fanciful characters interacting in detailed environments, rendered with black ink. An AI user fed Gi's portfolio into Stable Diffusion and used it to create a 'custom' model that made outputs in Gi's style. It had only been days after Gi died at 47 years old, and artists felt it was a gross violation

for AI users to view him as essentially a 'style font' to apply to outputs and consume and repurpose. [41]

The ease with which AI users can 'fine-tune' models to mimic or edit the works of specific artists is why it is essential for protections to be established to deny the use of artists' works as inputs in AI. All creators should be eligible for these protections, as relatively unknown creators can be impacted just as egregiously as ones with established careers and reputations, if not more so.

Protections that should be considered:

-AI models should be prohibited from 'training' on the works of artists that do not give explicit and affirmative consent. If the artists' works are not present in their training datasets, AI models would not have the visual, textual, or audial library to successfully mimic a specific artist at the onset.

-AI models should not allow inputs from end users and third parties. User-uploaded 'reference' material would allow for AI users to specifically 'fine-tune' models to mimic or alter specific artists' works and circumvent licenses and consent. If inputs are allowed, they should be vetted under the utmost scrutiny for proper licensing prior to use and a record should be kept so that there is evidence that artists' works were intentionally fed into the system.

-AI should be prohibited from accepting the names of characters, brands, and individuals as part of 'prompts'. For instance, an AI user could use 'mermaid' but not 'Ariel from *The Little Mermaid* by Disney' or 'Mermaid by -artist name-'. Preventing AI users from referencing names and titles, like those of artists or Disney and their copyrighted rendition of a mermaid, would ensure that the generator is not being called upon to specifically mimic a particular creator's expression.

As such, prompts should be catalogued and perhaps even present in each AI image's metadata so that copyright holders can act against the willful infringement of their intellectual property.

Ideally, however, artists' works and characters that are intellectual property should not be present at all in the datasets used in AI models, thus making the models incapable of producing accurate facsimiles even when instructed.

[40] Heikkilä, Melissa. "This artist is dominating AI generated art. And he's not happy about it." *MIT Technology Review.* September 16, 2022. www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/

[7] Nuttall, Jeremy. "Whose art is this, really? Inside Canadian artists' fight against AI". *Toronto Star.* February 2, 2023. www.thestar.com/news/canada/whose-art-is-this-really-inside-canadian-artists-fight-against-ai/article_54b0cb5c-7d67-5663-a46a-650b462da1ad.html

[41] Kim, Leo. "Korean Illustrator Kim Jung Gi's 'Resurrection' via AI Image Generator Is Orientalism in New Clothing". *ARTnews.* December 9, 2022. www.artnews.com/art-news/news/kim-jung-gi-death-stable-diffusion-artificial-intelligence-1234649787/

33:

114(b) of the Copyright Act should be granted attention in regards to AI, as it addresses copyright on broadcast transmissions, including live audio. It is possible that AI developers will use recordings of individuals from live events and other transmissions to circumvent copyright and publicity laws that would prohibit them from using an individual's voice or music as training data.

There is also the concern that devices in US citizens' homes (like smart devices capable of recording voices) will be used to harvest additional training data for AI applications. AI developers and companies could potentially claim these recordings as their own, and therefore free to use in training AI. Even if a consumer of these devices consents to their voice being used to train AI, those who visit their home or are in their public vicinity may be unaware of the devices' existence, and therefore cannot consent to the recording and use of their voice.

Regardless of the owner of the copyright of a particular recording- if a recording or live broadcast or transmission is made with the voice or musical performance of an individual, their license should be sought for the use of their likeness and work.

34:

Issues that the Copyright Office should study:

I believe it would be insightful for the Office inspect the contents of publicly accessible datasets commonly used by generative AI developers, like LAION-5B, its smaller subsets, and Books3 and its predecessors. These datasets are currently the focus of aforementioned cases instigated by authors and artists who found examples of their works within the datasets. The Copyright Office could organize a group that can compare the contents of these datasets with works registered through the Office. This way, the Office can see for itself the extent of registered works present in the datasets, and perhaps approach copyright holders about their views regarding the use of their work in generative AI.

Such questions to copyright holders could be-

-Have they been forced to find other work or are they preparing to change their career path because of generative AI? How has AI impacted the outlook they have for their career?

-Have they noticed a drop in engagement with their online presence, sales, or business due to competition from AI generated media?

-Have they discovered instances of AI models and users utilizing or specifically emulating their work or likeness?

-Would they ever voluntarily offer their works to AI developers for training, with or without compensation? If they want compensation, what would be equitable?

-Would they refuse to license their work for AI training, regardless of the offer?

-Have copyright holders seen a rise in contracts and terms with publishers, platforms, and distributors that would directly or vaguely give them license to use their works to train generative AI?

-How has scraping impacting copyright holders who use social media or are website owners, like independent artists and those who specialize in news reporting? Has scraping for AI

training resulted in copyright holders paywalling, deleting, privatizing, or withholding access to their media?

Additionally, the Copyright Office can study whether or not there have been 'scraping' efforts on the part of crawlers visiting the copyright registry itself- crawlers potentially belonging to AI companies. The Office may need to look into guidance along with the FTC and other organizations to assess the need for restrictions and consequences for scraping or collecting public, private, and copyrighted data for use in AI training- so as to protect copyright holders, consumers, and all Americans from having their works and data appropriated.

I would also suggest the Office conduct studies to better understand machine learning and how generative AI models' use of copyrighted media exploits artists and infringes upon their work. They should be demonstrated by researchers that do not work at the behest of corporations involved in commercial generative AI development, like Ben Zhao and his team at the University of Chicago. [25]

-

Once again, thank you for giving artists and performers like myself an opportunity to comment on the future of copyright, data privacy, and publicity rights, as well as the harms posed by generative AI.

-Chelle E.G.

10/30/2023

List of Cited Papers and Online Articles with Links:

[1] Harry H. Jiang, Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. "AI Art and its Impact on Artists". In *AAAI/ACM Conference on AI, Ethics, in Society (AIES'23), August 08--10, 2023, Montreal, QC, Canada.* AMN New York, NY, USA 12 Pages. doi.org/10.1145/3600211.3604681

[2] Whateley, Dan. "MrBeast Calls Out Tiktok for Allowing a Deepfake Version of Him in Ad" *Business Insider.* October 3, 2023 www.businessinsider.com/tiktok-ran-deepfake-ad-mrbeast-as-ai-generated-content-spreads-2023-10

[3] Britton, Bianca. "They Appeared in Deepfake Porn Videos Without their Consent. Few Laws Protect Them." *NBC News.* February 14, 2023. www.nbcnews.com/tech/internet/deepfake-twitch-porn-atrioc-qtcinderella-maya-higa-pokimane-rcna69372

[4] Duffy, Clare. "An Author Says AI is 'Writing' Unauthorized Books Being Sold Under Her Name on Amazon"- *CNN.* Aug 10, 2023 www.cnn.com/2023/08/10/tech/ai-generated-books-amazon/

[5] Tiffany, Kaitlyn. "AI-Generated Junk Is Flooding Etsy"- *The Atlantic.* June 15, 2023 www.theatlantic.com/technology/archive/2023/06/ai-chatgpt-side-hustle/674415/

[6] Peterson, Andrea. "SAG-AFTRA is fighting a dystopian AI takeover so you don't have to." *MSNBC.* July 18, 2023. www.msnbc.com/opinion/msnbc-opinion/wga-sag-aftra-strike-writers-actors-ai-hollywood-rcna94318

[7] Nuttall, Jeremy. "Whose art is this, really? Inside Canadian artists' fight against AI". *Toronto Star.* February 2, 2023. www.thestar.com/news/canada/whose-art-is-this-really-inside-canadian-artists-fight-against-ai/article_54b0cb5c-7d67-5663-a46a-650b462da1ad.html

[8] Gordon, Anna and Perrigo, Billy. "E.U Takes a Step Closer to Passing the World's Most Comprehensive AI Regulation." *TIME.* July 14, 2023. time.com/6287136/eu-ai-regulation/

[9] O'Carroll, Lisa. "EU 'in touching distance' of world's first laws regulating artificial intelligence". *The Guardian.* October 24, 2023. www.theguardian.com/technology/2023/oct/24/eu-touching-distance-world-first-law-regulating-artificial-intelligence-dragos-tudorache

[10] Creamer, Ella. "MPs criticise UK government's handling of copyright policy related to AI". *The Guardian.* August 30, 2023. www.theguardian.com/books/2023/aug/30/mps-criticise-uk-governments-handling-of-copyright-policy-related-to-ai

[11] Yang, Zeyi. "China isn't waiting to set down rules on generative AI". *MIT Technology Review.* May 31, 2023.
www.technologyreview.com/2023/05/31/1073743/china-generative-ai-quick-regulation/

[12] Paul, Katie. "Meta's new AI assistant trained on public Facebook and Instagram posts". *Reuters.* September 28, 2023.
www.reuters.com/technology/metas-new-ai-chatbot-trained-public-facebook-instagram-posts-2023-09-28/

[13] Milmo, Dan. "Mushroom pickers urged to avoid foraging books on Amazon that appear to be written by AI". *The Guardian.* September 1, 2023.
www.theguardian.com/technology/2023/sep/01/mushroom-pickers-urged-to-avoid-foraging-books-on-amazon-that-appear-to-be-written-by-ai?ref=distilleddaily.news

[14] Lazarro, Sage. "The A.I. industry confronts life after data scraping". *Fortune.* August 29, 2023.
fortune.com/2023/08/29/the-a-i-industry-confronts-life-after-data-scraping/

[15] Rowe, Niamh. "'It's destroyed me completely': Kenyan moderators decry toll of training of AI models". *The Guardian*. August 2, 2023.
www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai

[16] Javaid, Maham. "The magic Avatars you paid $3.99 for is probably stolen, artists say." The *Washington Post.* December 9, 2022.
www.washingtonpost.com/technology/2022/12/09/lensa-apps-magic-avatars-ai-stolen-data-compromised-ethics/

[17] Leffer, Lauren. "Your Personal Information is Probably Being Used to Train Generative AI Models". *The Scientific American*. October 19, 2023.
www.scientificamerican.com/article/your-personal-information-is-probably-being-used-to-train-generative-ai-models/

[18] Asmelash, Leah. "These books are being used to train A.I. No one told the authors." *CNN.* October 8, 2023.
edition.cnn.com/2023/10/08/style/ai-books3-authors-nora-roberts-cec/

[19] Goldman, Sharon. "Adobe Stock creators aren't happy with Firefly, the company's 'commercially safe' gen AI tool". *VentureBeat.* June 20, 2023.
venturebeat.com/ai/adobe-stock-creators-arent-happy-with-firefly-the-companys-commercially-safe-gen-ai-tool/

[20] Lardinois, Frederic. "Adobe Firefly can now generate more realistic images". *TechCrunch.* October 10, 2023.
techcrunch.com/2023/10/10/adobe-firefly-can-now-generate-more-realistic-images/

[21] John, Daniel. "Artists complain of AI 'copyright infringement' on Adobe Stock". *Creative Bloq*. August 24, 2023.
www.creativebloq.com/news/adobe-copyright-ai

[22] Xiang, Chloe. "A Photographer Tried to Get His Photos Removed from an AI Dataset. He Got an Invoice Instead." *Vice*. April 28, 2023.
www.vice.com/en/article/pkapb7/a-photographer-tried-to-get-his-photos-removed-from-an-ai-dataset-he-got-an-invoice-instead

[23] Xiang, Chloe. "AI Spits Out Exact Copies of Training Images, Real People, Logos, Researchers Find". *Vice.* February 1, 2023.
www.vice.com/en/article/m7gznn/ai-spits-out-exact-copies-of-training-images-real-people-logos-researchers-find

[24] Harrington, Kayla. "An AI Program Imagined Barbies From Around the World and Wouldn't You Know It, They Turned Out Super Racist". *The Mary Sue.* August 4, 2023.
www.themarysue.com/ai-created-racially-stereotyped-barbies-from-around-the-world/

[25] Heikkilä, Melissa. "This new data poisoning tool lets artists fight back against generative AI". *MIT Technology Review.* October 23, 2023.
www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/

[26] Pastis, Stephen. "A.I.'s un-learning problem: Researchers say it's virtually impossible to make an A.I. model 'forget' the things is learns from private user data." *Fortune.* August 30, 2023.
fortune.com/2023/08/30/researchers-impossible-remove-private-user-data-delete-trained-ai-models/

[27] Carlisle, Madeleine. "How Google's Big Supreme Court Victory Could Change Software Forever". *TIME.* April 6, 2021.
time.com/5952718/google-oracle-supreme-court/

[28] Bandara, Pesala. "Nature Photographer Who Sued Andy Warhol Over Her Photo of Flowers Dies". *Peta Pixel.* September 11, 2023.
petapixel.com/2023/09/11/nature-photographer-who-sued-andy-warhol-over-her-photo-of-flowers-dies/

[29] Ducharme, Jamie. "Can AI Chatbots Ever Replace Human Therapists?". *TIME.* October 4, 2023.
time.com/6320378/ai-therapy-chatbots/

[30] "From ChatGPT to Deepfake Apps: A Running List of AI Lawsuits". *The Fashion Law*. October 19, 2023.
www.thefashionlaw.com/from-chatgpt-to-deepfake-creating-apps-a-running-list-of-key-ai-lawsuits/

[31] Sheng, Ellen. "In generative AI legal Wild West, the courtroom battles are just getting started". *CNBC.* April 3, 2023.
www.cnbc.com/2023/04/03/in-generative-ai-legal-wild-west-lawsuits-are-just-getting-started.html

[32] Foley, Joseph. "DeviantArt breaks silence over fierce backlash for its new AI art tool". *Creative Bloq*. November 15, 2022.
www.creativebloq.com/news/deviantart-dreamup-ai

[33] Wiggers, Kyle. "How much can artists make from generative AI? Vendors won't say". *TechCrunch.* September 30, 2023.
techcrunch.com/2023/09/30/how-much-can-artists-make-from-generative-ai-vendors-wont-say/

[34] Tidy, Joe. "Paedophiles using AI to turn singers and film stars into kids". *BBC.* October 24, 2023.
www.bbc.co.uk/news/technology-67172231

[35] Claburn, Thomas. "Big Tech loves talking up privacy- while trying to kill privacy legislation". *The Register.* May 27, 2022.
www.theregister.com/2022/05/27/big_tech_privacy/

[36] Setty, Riddhi. "Getty Images Sues Stability AI Over Art Generator IP Violations". *Bloomberg Law.* February 6, 2023.
news.bloomberglaw.com/ip-law/getty-images-sues-stability-ai-over-art-generator-ip-violations

[37] Murphy Kelly, Samantha. "ChatGPT creator pulls AI detection tool due to 'low rate of accuracy'". *CNN.* July 25, 2023.
edition.cnn.com/2023/07/25/tech/openai-ai-detection-tool/

[38] Steinbuch, Yaron. "Traumatized Ariz. mom recalls sick AI kidnapping scam in gripping testimony to Congress". *New York Post.* June 14, 2023.)
nypost.com/2023/06/14/ariz-mom-recalls-sick-ai-scam-in-gripping-testimony-to-congress/

[39] Tenbarge, Kat. "Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy." *NBC News.* March 27, 2023.
www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071

[40] Heikkilä, Melissa. "This artist is dominating AI generated art. And he's not happy about it." *MIT Technology Review.* September 16, 2022.
www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/

[41] Kim, Leo. "Korean Illustrator Kim Jung Gi's 'Resurrection' via AI Image Generator Is Orientalism in New Clothing". *ARTnews.* December 9, 2022.
www.artnews.com/art-news/news/kim-jung-gi-death-stable-diffusion-artificial-intelligence-1234649787/