

COMMENTS OF PROFESSOR SANDRA AISTARS

George Mason University, Antonin Scalia Law School

I thank the U.S. Copyright Office (the “Office” or the “USCO”) for the opportunity to submit comments in response to the [notice of inquiry and request for comments](#) (“NOI”) published by the Office in the Federal Register on August 30, 2023 (and [the extended comment deadline](#) noticed on September 21, 2023). It is appropriate that the Office is studying the copyright law and policy issues raised by generative artificial intelligence (“GAI”) systems.

I am a member of the Copyright Alliance Academic Advisory Board and AI working committee. I join in the comments submitted by the Copyright Alliance. I submit these additional brief comments to add a few personal notes to the record.

I am a professor and a practitioner of law who has worked at the cross-section of art and technology for more than twenty-five years. I am the founder and director of the Arts & Entertainment Advocacy Clinic at George Mason University’s Scalia Law School, where I supervise students delivering pro bono copyright counseling services to individuals and small businesses in the arts. I am a full-time professor at the law school, and I teach various doctrinal classes including in entertainment law, ethics (professional responsibility) and other subjects. I am also a Senior Scholar and the Senior Fellow for Copyright Policy and Research at the school’s Center for Intellectual Property x Innovation Policy. Along with my responsibilities to the school, I currently serve on the Board of Directors of the Washington Area Lawyers for the Arts. I also work closely with the Institute for IP and Social Justice as a standing member of its planning committee. Although my work is currently focused on academia and on serving the community, I began my career as an attorney at the law firm Weil, Gotshal and Manges LLP, representing technology

companies including in negotiating, and later navigating and developing successful consumer facing business models in compliance with, the copyright obligations imposed by the Digital Millennium Copyright Act. I later held various roles at the media company Time Warner Inc., ending as Vice President and Associate General Counsel for Intellectual Property for the corporation. Immediately before transitioning to academia, I was CEO of the Copyright Alliance -- a non-profit, non-partisan public interest and educational organization that represents the interests of a vast swath of the professional creative community with copyright interests. At various times in my career, I have chaired cross-industry technology development and/or standards efforts to voluntarily develop means to comply with aspects of copyright law for the benefit of users, consumers and innovators alike.

I mention my professional associations for identification purposes only. I have no personal stake or interest in the outcome of these proceedings other than an interest in ensuring that copyright law and regulations are interpreted and continue to develop in a manner consistent with its constitutional and statutory foundations so that creativity and innovation will continue to flourish.¹

Introductory Statement

The Office's request for comments is phrased to elicit factual reports and views based on experience with the operation of AI systems and technologies.² My experience comes from my years of private practice representing clients across the spectrum of creativity and innovation including major technology companies, international media companies, copyright owners from all creative disciplines, and from representing individual artists and small businesses who I not only currently advocate for through the Mason Arts and Entertainment Advocacy Clinic, but whose interests I have also represented throughout my career on a pro bono basis. My comments are also informed by a current research project in which I am surveying artists from various artistic disciplines who already use or have experimented with using GAI tools in their professional

¹ In the interest of full disclosure, the Mason Arts and Entertainment Advocacy Clinic will be co-presenting an educational program on current legal and regulatory developments relevant to the use of Artificial Intelligence Systems for artists on Nov. 14, 2023 together with Washington Area Lawyers for the Arts and National Association of Voice Actors. Neither I nor the clinic currently represent either of these entities in these or any other proceedings related to the regulation of AI systems.

² 88 Fed. Reg. Notice, 4 (2023) (The United States Copyright Office seeks comments on legal and ethical issues that may arise from the application of copyright law to AI systems).

creative work. My comments also reflect on a recent incident in which my own scholarly work was purportedly summarized by an AI chatbot and reported to a colleague, unprompted, prior to the work's publication.

GAI tools have great potential. As so eloquently elaborated in the Copyright Alliance filing, GAI tools have immense potential to unlock new creative expression for both professional and amateur users in a variety of fields, and to allow authors to shape, channel and mold their own voices applying GAI systems in ways that will bring untold new knowledge and original expression to the public. Artists and creators from across the creative disciplines have been using GAI tools already, and many are eager to use them in the future. Examples of uses include:

- Overcoming physical challenges that hinder artists' full engagement in their chosen creative discipline
- Ideation, and development of non-protected elements of literary or other works
- Automation of repetitive or time-consuming tasks
- Recombination or manipulation of one's own works with the assistance of AI tools
- Ethical use of an artist's personal data to create synthetic assets the artist licenses to expand their personal reach (e.g. a voice actor offering a limited license for a synthetic voice while being able to simultaneously live act in another project).

The examples listed in the Copyright Alliance filing demonstrate some of the new ways the creative sector is using GAI technology. However, there can be hesitancy – especially among professional authors of copyrighted works -- to fully engage the available tools and deploy them to create professional work on a meaningful scale. While the Copyright Office, other agencies and Congress have hosted many listening sessions and hearings for parties to put their views and opinions on the public record, in many discussions — both public and private — many of the same questions, uncertainties, and fears continue to arise. At the moment, there appear to be conflicting views (even about the basics of what GAI technologies do), but little certainty. Therefore, many participants who might otherwise embrace exciting technological developments for the benefit of their creative work hang back.

Some artists are taking a wait and see attitude about GAI tools due to legal uncertainties.

GAI tools could be transformational creative and scientific tools for human advancement if they were designed and implemented in a transparent or collaborative fashion. Artists and authors are interested in tools that ethically extend and empower artistic and scientific vision. They want to

understand the implications of using the tools, and how they can more fully exercise their creative voices when doing so. At the moment, however, it is not evident that designers of such tools are inclined to fully partner with their potential users in the creative and scientific communities in order to ethically fulfill those visions. Instead, many tools are being delivered to the marketplace in a less than transparent fashion by companies reluctant to engage in dialogue with professional artists or scientists who could be their greatest users. As a result, many GAI tools are drawing criticism and even lawsuits from the creative community.

Transparency is key to building safe, ethical, and unbiased GAI systems that respect copyright.

Transparency is important to creators to allow them to mitigate their own risk of legal liability, and so they can make informed decisions about the commercial and societal goals they lend their names, images, likenesses, and art to. Artists are concerned about (1) the use of their copyrighted works without authorization and compensation to train GAI tools – such training requires the making of an independent database copy without authorization; (2) the use of GAI tools once trained to create infringing works that too closely resemble the author's input work, and (3) the potential community impact of the tools that make use of their works. Their concerns are not likely to be addressed by considering GAI tools in a sector-specific fashion. For instance, it is not readily apparent that an artist can prevent the use of her work to train AI tools that perform malicious activities or to assist in surveillance of the artist's community unless all GAI tools have to meet appropriate standards consistent with U.S. copyright law regardless of the domain in which the tool is destined to be used. Subjecting some tools to lower standards and others to higher standards regarding copyright exposes creators to the risk that infringing copies of their work will be used to train tools that are problematic.

Transparency should be combined with explainability, security and trust.

The GAI tools most widely commented on and criticized are general-purpose AI systems, trained on vast amounts of text and images scraped from the internet. Often the developers of such tools obscure the source of training materials for their GAI systems. Such systems are understandably causing consternation among authors who have reason to believe their works are being consumed for commercial use without consultation or compensation.

While GAI tools are feted for their ¹supposed autonomy, they are nevertheless designed by humans and reflect the choices and priorities of the people and companies that design them. In the creative industries AI systems can be used to automate tedious tasks, to overcome challenges, and to make predictions in line with their programming or to suggest improvements in works ways authors and innovators may find helpful to name just a few promising possibilities. However, GAI systems are not being deployed by companies offering the tools purely for altruistic goals – nor should they be.³ Instead, the benefits possible from using AI systems are always part of a broader overall project plan for the systems. Unfortunately, the intentions of AI tool designers are not necessarily apparent to the users of the tools. Indeed, a user may have little or no visibility into the business model behind the tool it interacts with, or even know whose tool the user is interacting with at all. It is even possible that some users cannot discern when or if they are interfacing with AI systems.

Unlike traditional artists' tools which specialized, for profit businesses historically developed to cater to a specific community's needs, AI tools appear — at least for the time being — to be emerging much like free search engines, user generated content sites and other generalized Web 2.0 tools of the past. GAI tools like Claude, Midjourney, Dall-e, and others are not being developed and marketed with the input and nurturing of the arts community², but rather are released as widely-available tools suitable for a wide variety of users whose engagement with the tools may be monetized by the AI developers on multiple levels. Tools may apply a variety of business models some of which are not visible to the end user. Business models can include licensing the technology to third party app developers and charging the apps a session based fee for engaging the tool, charging users session based fees, requiring users to purchase a subscription to a site or otherwise agree to join a community where the tools are available. All of these models may require the user to trade private data about themselves or how they use the tools or grant the GAI developer the right to use copyrighted works created with the tool³ even if other compensation is being paid

¹ Mark Andreessen, *The Techno-Optimist Manifesto* (Oct. 26, 2023) <https://a16z.com/the-techno-optimistmanifesto/> last visited October 30, 2023. (“people only do things for other people for three reasons – love, money or force. Love doesn’t scale, so the economy can only run on money or force. The force experiment has been run and found wanting. Let’s stick with money. We believe the ultimate moral defense of markets is that they divert people who otherwise would raise armies and start religions into peacefully productive pursuits.”)

² Darian Woods and Adrian Ma, *Artists File Class-Action Lawsuit Saying AI Artwork Violates Copyright Laws*, NPR (Feb. 3, 2023), <https://www.npr.org/2023/02/03/1154091957/artists-file-class-action-lawsuit-saying-ai-artworkviolates-copyright-laws>.

³ *Privacy Policy*, OpenAI (June 23, 2023), <https://openai.com/policies/privacy-policy>; *Terms of Service*, Midjourney (July 21, 2023), <https://docs.midjourney.com/docs/terms-of-service>.

to the GAI developer. Developers of AI services may be supported by advertising or may sell private data about users to security agencies, profilers, or other application builders that may use them for legitimate or for nefarious reasons.⁴

Copyright implications of AI tools cannot be understood in isolation. It is important to understand the technical and business choices developers make not only about copyright issues, but about overall design and project planning/problem formulation for the AI system in order to understand how authors' copyright choices are affected. Technical and business decisions mentioned are significant for both the authors whose works are originally used to train the tools, and for later authors who become users of the tools (whether or not the works produced are ultimately copyrightable).

Copyright (and other) harms must be examined both from the perspective of the author whose work(s) are ingested into a tool for training and possible use in outputs, and from the perspective of the user of the tool, who may or may not produce a copyrightable work. Strategic decisions made by project designers in setting the parameters and goals of a project can result in harm not only to the creators whose works might be ingested/output and thus infringed by the AI tool, but to users of the AI systems who may themselves become authors of copyrightable works while using the tools. Harms to users of GAI tools could include copyright related harms such as the infringement of their own exclusive rights, but also privacy and right of publicity related harms related to any data collected by the AI system while the person uses the system, and how that data is treated. Such harms could implicate users' interests in maintaining human dignity, equality, privacy, and the rule of law generally – just like those harms are implicated when the copyrighted work or the identities of rights holders are ingested into GAI tools and used to train them in the first instance. Therefore, while conducting this inquiry the Copyright Office should not lose sight of the overall project parameter for the training and use of a given GAI technology model.

Humans are fallible, and input from outside perspectives can be helpful. Even with the best of intentions it is possible for a product designer to fail to anticipate technical and social failure

⁴ Cameron F. Kerry, *Protecting Privacy in an AI-Driven World*, Brookings Institution (Feb. 10, 2020), <https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/> (“Much of the most privacy-sensitive data analysis today—such as search algorithms, recommendation engines, and adtech networks—are driven by machine learning and decisions by algorithms. As artificial intelligence evolves, it magnifies the ability to use personal information in ways that can intrude on privacy interests by raising analysis of personal information to new levels of power and speed.”).

modes. For instance, a technology may not work as intended upon alpha and beta testing or fail to meet certain product development criteria upon use. Alternatively, a technology may cause injustices or harms to users that may not have been intended or understood when the product was being designed. Effective standards and certification/registration testing should aim to surface both. Technology designers just like authors of copyrighted works are understandably sensitive to confidentiality and other concerns that arise when disclosing project ideas for community feedback or audit. Likewise, government pre-approvals or licensing of AI systems are likely to meet with objections either as a whole or when details of proposed regulations are probed. Nevertheless, responsible innovative companies are constantly testing and updating products to address changes in the global regulatory environment and to deter malicious actors. It is not unusual to submit products for internal or third- party compliance testing to ensure that the product or service meets all applicable regulatory and legal requirements. In some product areas like green-design, companies and designers can submit products and services to obtain voluntary acknowledgments that they meet entirely voluntary design standards in order to send particular market signals. Unfortunately, without any actionable USCO guidance concerning GAI product design (as opposed to guidance concerning registration requirements for works authored with the assistance of GAI tools) there is no clear guidance for GAI companies who wish to incorporate reliable copyright standards in their testing regimes.

The Office should consider offering formal guidance on copyright standards all GAI tools should meet. It could additionally offer voluntary certification and/or registration by the Office of such tools.

Some comments on history. The history and adoption of technologies like photography have been put forward by some as a model for analyzing current topics related to AI tools.⁵ Some

⁵ *R Street Signs Coalition Letter Expressing Concerns With Future AI Copyright Legislation* (Sep. 11, 2023) <https://www.rstreet.org/outreach/r-street-signs-coalition-letter-expressing-concerns-with-future-ai-copyrightlegislation/>

commentators argue that in the 1800s “artists objected to the invention of photography, arguing that it would render the paintbrush obsolete,” and that concerns from creators that GAI systems do not comply with copyright laws are the same. Such rhetoric misses the mark.

Photography did not depend on the reproduction of any copyrighted works for its existence. Not so with GAI. In order to function at all a GAI system must first be trained on large databases of works which invariably contain copyrighted content. Because photography does not require the reproduction of copyrighted works in order to enable the technology, the risk/use models for photography can be assessed *ex post facto*, by applying either originality/authorship analysis or copyright infringement analysis to the photographs at issue as appropriate. In contrast, addressing only the outputs of GAI technology does not resolve concerns about infringements occurring in training GAI systems. There are two separate infringement points and reproductions occurring, each of which is a separate actionable infringement of the copyright owners’ right to control the reproduction of the work under section 106(1).

In the case of photography, in the rare case that an image fixed by photography does not capture the original authorship of the user of the technology, the image can be denied protection of copyright. Unfortunately, *ex post facto* approaches addressed only to the outputs of a GAI device do not solve problems stemming from unlicensed inputs because the database copy used for ingestion and training a GAI system is a separate infringing copy. A remedy addressed only to the outputs of a GAI is ineffective to cure harm to the author whose work was used to train the GAI system, because the end user of the system may or may not seek to register a work resulting from interaction with the GAI, and may not ultimately care if the work is denied protection. As for the GAI developer, the developer may ultimately benefit if works created using their tool are not copyrightable, because it limits any licensing obligations the GAI developer might otherwise have to the users of its tool regarding works they create using the tool. Such works can be indiscriminately swept back in and used to further train the GAI system or sold/licensed to others without any concerns of violating the GAI user’s copyright interests.

The Office should provide Guidance to GAI developers on applicable copyright standards. These comments are not intended to suggest that the Office step back from its legislatively mandated role to regulate in this area. Quite to the contrary, I write to underscore the importance of the Office's work, and to urge the Office to consider reframing the inquiry more precisely and effectively. Rather than focusing its AI Initiative mainly on the Office's registration function regarding expanding or contracting the possibility of registration of works generated by AIs, or works containing material generated by AIs, (a question which standing alone cannot fully address any of the infringement challenges associated with ingestion of works or speak to privacy or other risks identified with AI systems generally) the Office should concentrate its attention more directly on providing guidance on how GAI systems can meet their copyright obligations. This could include establishing a voluntary certification or registration option for those who choose to use it. A possible end result of the inquiry could be

- a new copyright compliance tool guidance document for GAI developers and
- a tool for authors and users of GAI tools to help them understand, evaluate, interrogate, and pose questions about the copyright and other implications raised by the operation of the tools.
- A voluntary certification/registration system for GAI tools to ensure they meet certain baseline copyright standards in a secure fashion.

Ongoing community input facilitated by the above-suggested tools and guidance documents could be helpful in supporting any new voluntary registration or certification system for GAI tools whereby GAI tools would be encouraged to seek certification against a set of appropriate copyright and related standards to ensure that they are developed in a manner that sufficiently reflects the circumstances, interests, needs and priorities of the communities that will use them, and ensures certain ethical algorithmic standards are implemented in the design and use of systems.⁶ Such an approach might better address or at least supplement the copyright systems' interests in limiting infringements than efforts that either grant or deny registration to works created by registrants using GAI systems.

⁶ See Peter Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 Florida Law Rev., 331, 370 (2020) ("Even though trusting designers to come up with best practices and codes of conduct can be highly promising, there also needs to be external responses. At the domestic level, laws should be put in place to ensure

Clarity on standards will mitigate concerns about litigation risk. It is also important to consider putting such a guidance and certification system in place in order to encourage authors to use GAI tools to enhance their creative reach. At the moment, discussions with authors in various disciplines often reveal a reluctance to use AI tools in professional work or to even admit to casual experimentation with GAI tools for fear of the consequences use of such tools may pose. Those risks include limits on claims of authorship for work generated using GAI tools, as well as risks associated with infringement. The Office's current registration guidance also raises uncertainties about penalties associated with registration attempts.

For instance, litigation related risks can include the following:

- a user of a GAI tool may doubt whether they will be able represent and warrant that they are owners of all the IP in works they create using GAI tools.
- Their worries can actually increase rather than be relieved by recent offers by some GAI tool proponents to indemnify users against lawsuits that might arise from the use of their tools. Such offers may be interpreted as looking for a good test case, and as indications that there are indeed significant risks with using the tools.
- It may be especially problematic if an author licenses and indemnifies others against claims involving work they create using GAI tools.
- Users of GAI tools may likewise fear that if there is ever litigation over their works, they will not be able to rely on a presumption of validity for copyright registration certificates issued by the Office and that instead copyrightability will become a question to be litigated in every dispute – prolonging litigation and making it more expensive.
- Even worse, due to lack of transparency from designers of GAI tools about the design and operation of GAI tools, authors may worry they may be asked to produce evidence concerning the contributions of the GAI tool versus the contributions of the author themselves which the author will not be able to procure.
- All of these concerns are amplified if the author is asked to make representations to the Office or to another party about their works during the registration process – especially if the representations must be made on forms that are not easily parsed or applied to their works, come with warnings that they are made under greater ethical standards in the design and use of algorithms, At the international level, treaties, soft law recommendations, or

other normative documents can be utilized to help facilitate the development of internationally acceptable standards.”) penalty of perjury⁷⁸ and/or with risk of loss of registration status¹⁰ or if other penalties for the failure to be fully transparent apply to the author.¹¹

- There is an additional risk that the lack of transparency by GAI tool designers will require a copyright registrant and the Office each to marshal their own facts, independent expert opinions and analysis about the operation of tools when seeking registration of a work created using GAI tools as part of the process⁹ instead of referencing a central fact record provided by the developer of the tool themselves. Individualized, factbased inquiries into registrations in order to serve a larger mission aimed at regulating the GAI tool itself in some fashion seem unlikely to be practical on a large scale or to ultimately serve the interests of the Office or of registrants or GAI systems in generating accurate and predictable results for all concerned parties.

General Questions:

1. As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

⁷ See *Form VA*, United States Copyright Office 4 (2023), <https://www.copyright.gov/forms/formva.pdf> (“17 U.S.C. §506(e): Any person who knowingly makes a false representation of a material fact in the application for copyright registration provided for by section 409, or in any written statement filed in connection with the application, shall be fined not more than \$2,500.”).

⁸ Fed. Reg. Notice, 4 (2023) (“Applicants have a duty to disclose the inclusion of AI-generated content in a work submitted for registration and to provide a brief explanation of the human author’s contributions to the work.”).¹¹ *Id.* (“Applicants who fail to update the public record after obtaining a registration for material generated by AI risk losing the benefits of the registration.”).

⁹ *Re: Zarya of the Dawn (Registration # VAu001480196)*, United States Copyright Office, 5 (Feb. 21, 2023) (“The Office’s understanding is based on the letter’s description of the artificial intelligence service, the Office’s own knowledge, and Midjourney’s public documentation, of which the Office takes administrative notice.”).

As described in the introduction, GAI tools have vast potential to inform and entertain, as well as to assist humans in their lives in myriad ways. They can provide company, generate ideas and engage in conversation, synthesize information, produce images, text, music, and more. These are merely high-level descriptions of some of the possible functions GAI tools can perform. The benefits and risks of the technology will be determined by the innovators who develop the product parameters for the tools and set them in motion, and the users who engage them – either intentionally or not. However, focusing on whether the material output by the generative AI should be copyrightable does not provide a satisfactory solution to the risks thus far identified by various authors of copyrighted works who assert their works have been infringed by AI systems.

Examples of lawsuits currently pending against AI developers

The Office is no doubt well aware of the numerous lawsuits filed by workers, authors and business in the creative sector identifying some of the risks of AI technology. Those lawsuits emphasize (1) concerns with ingestion and copying of works for training purposes; (2) infringements as a result of prompting AI systems to create new works; and (3) the outputs of AI systems and whether or not those outputs are substantially similar to complaining copyright owners' works. The Copyright Alliance filing contains an exhibit summarizing numerous of these suits.

Example of other issues raised by AI systems

One of the ways companies are using AI tools that rely on copyrighted works is to generate commercial messages to users and to engage them in discussion with chatbots. These types of uses can present novel factual scenarios that could meld together copyright concerns, right of publicity and trademark concerns and privacy issues. A recent feature section in the New York Times on AI and Chatbots describes the many ways that AI powered chatbots trained on Large Language Learning Models are being deployed. As is the case with other GAI systems, these systems depend on large troves of copyrightable works for their training corpus.¹⁰ Some companies have revealed the sources of their training materials,

¹⁰ Mike Isaac and Cade Metz, *Meet the AI Jane Austen: Meta Weaves A.I. Throughout Its Apps*, The New York Times (Sept. 27, 2023), <https://www.nytimes.com/2023/09/27/technology/meta-ai-celebrities.html>.

others are less transparent. Some of the popularly used training sets include The Pile, Books 3, LAION, and WebVid-10M. Investigative reporters have uncovered that these datasets include pirated copies of entire copyrighted works.¹¹

Companies like Meta, OpenAI, Google, Microsoft and others are deploying GAI that can generate text, images and other media on their own across their apps.¹² The companies are using the technology to draw consumers to their apps – which may rely on advertising for revenue – in order to increase consumer engagement.¹⁶ The more time a user spends in a given app, the more ads the AI or app developer can expose the user to. Thus, even where an app is not being used to create material from the output of the AI that is substantially similar to the training data the app is trained from, some lawsuits assert that the wholesale copying of works for a commercial purpose to facilitate the operation of the app should nevertheless be compensable to the rightsholder.¹³

A New York Times article describes one way a GAI chatbot could interact with users to generate commercial benefits for an app. Two friends chatting on What's App could turn to Meta's chatbot Meta AI to ask for a recommendation of a restaurant. The app would mine Microsoft's Bing search engine for search results and return a short conversational response. If the friends wanted more information, they could click on linked footnotes in the response that would redirect them to Bing search results in a different window.¹⁴ Although *Authors' Guild v Google Books*¹⁵ has been suggested by some as applicable to the analysis, courts have confirmed that *Google Books* is a narrow ruling which should be understood in the context of its own facts.¹⁶ In *Google Books* although the copying of a large corpus of books from libraries was found to be a fair use, the copying was done to

¹¹ Alex Reisner, *Revealed: The Authors Whose Pirated Books Are Powering Generative AI*, The Atlantic (Aug. 19, 2023), <https://www.theatlantic.com/technology/archive/2023/08/books3-ai-meta-llama-pirated-books/675063/>.

¹² *Id.*¹⁶

Id.

¹³ Alexandra Alter and Elizabeth A. Harris, *Franzen, Grisham and Other Prominent Authors Sue OpenAI*, The New York Times (Sept. 20, 2023), <https://www.nytimes.com/2023/09/20/books/authors-openai-lawsuit-chatgptcopyright.html>.

¹⁴ *Id.*

¹⁵ *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015).

¹⁶ *Hachette Book Group, Inc. v. Internet Archive*, 2023 WL 2623787 (S.D.N.Y. 2023).²¹ *Authors Guild*, 804 F.3d at 209.

facilitate a limited purpose – snippet view – that did not compete with the interests of the authors, and in fact was arguably intended to facilitate discovery and purchase of their copyrighted works.²¹ Moreover, the copies of the works were encrypted and secured to the Google platform to prevent leakage and use by others,¹⁷ in contrast with GAI training sets which are often shared among multiple services and can even be sourced from infringing or laundered data.¹⁸ It is thus not evident that courts would approve the ingestion of the corpus of copied training data as a fair use given the dubious origins of data (which in itself should disqualify the corpus from attaining fair use status)¹⁹, the commercial use of the data, and the repurposing of the works that is well beyond the narrow approach found permissible in *Google Books*.

Unexpected interactions with chatbots. Not all chatbot interactions will be innocuous. In instances where a chatbot behaves in a questionable or problematic fashion copyright owners may be troubled that training on their work is enabling such behavior. Examples could include chatbots that engage in offensive or discriminatory behavior or perpetuate lies, and misinformation.

Other interactions can simply be hard to understand and can therefore raise privacy and related worries. Our academic center experienced an unexpected interaction with the Bing search engine and its related chatbot in which the chatbot appeared to be pushing an unknown user's or an autonomously generated chat to us as a news item. This interaction raised a variety of questions for us.

A colleague who has a news alert for my name received a notification mentioning me. When he clicked on the news alert a chat opened with a query that was already populated. It asked “Substantial similarity is a key concept in copyright law. What is the Aistars test

¹⁷ *Id.* at 228.

¹⁸ Alex Reisner, *Revealed: The Authors Whose Pirated Books Are Powering Generative AI*, The Atlantic (Aug. 19, 2023), <https://www.theatlantic.com/technology/archive/2023/08/books3-ai-meta-llama-pirated-books/675063/>.

¹⁹ *Can I Use Someone Else's Work? Can Someone Else Use Mine?* United States Copyright Office, <https://www.copyright.gov/help/faq/faq-fairuse.html#:~:text=There%20are%20circumstances%20under%20the,movies%20in%20the%20public%20domain%3F>.

for determining whether two works are substantially similar? How has this test been applied in recent cases?” The response from the chatbot stated that a test named after me exists, and described a test I have never suggested.

The chat/news result seems problematic for several reasons.

- *The summary of the article is inaccurate.* There is no such thing as an “Aistars test” for substantial similarity (or any other legal concept) nor does the chat accurately reflect the content of the article. It is a well-known problem that GAI’s provide inaccurate summaries and other faulty information in response to user queries. However, such chats have in the past remained private to the user who is interacting with the chatbot. A user could use discretion and edit chat prompts or do further research to determine whether to rely on or otherwise publicize the results. Here the chatbot appears to be publishing queries as news items to other users.
- *Accessing and using the article to train an AI tool violates SSRN’s terms of service.* The article referenced in the chat is only legitimately available on SSRN – a website that requires a user account to access materials posted there. SSRN’s Terms of Service prohibit the commercial use of posted materials to train AI tools. The Terms of Service for SSRN state in relevant part:

You may not copy, display, distribute, modify, publish, reproduce, store, transmit, post, translate or create other derivative works (including resulting from the use of artificial intelligence tools) from, or sell, rent or license all or any part of the Content, or products or services obtained from the Services, in any medium to anyone, except as otherwise expressly permitted under these Terms and Conditions, or any relevant license or subscription agreement or authorization by us. . .

You may not use Content from the Services in combination with an artificial intelligence tool, (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool). You may not use any robots, spiders, crawlers or other automated downloading programs, algorithms or devices, or any similar or equivalent manual process, to: (i) continuously and automatically search, scrape, extract, deep link or index any Content; (ii) harvest personal information from the Services for purposes of sending unsolicited or unauthorized material; or (iii) cause disruption to the working of the Services or any other person’s use of the Services. If the Services contain robot exclusion files or robot exclusion headers, you agree to honor them and not use any

device, software or routine to bypass them. You may not attempt to gain unauthorized access to any portion or feature of the Services, any other systems or networks connected to the Services or to any Elsevier server, or any of the products or services provided on, accessed from or distributed through the Services. You may not probe, scan or test the vulnerability of the Services or any network connected to the Services or breach or attempt to breach the security or authentication measures on the Services or any network connected to the Services.

- *Misrepresentations of work can generally be expected to have a negative effect on readership and/or author interest in maintaining creative control of a project pre-publication.* There are many reasons why copyright law affords authors the first right of publication. In *Harper & Row v. Nation Enterprises*,²⁰ the Supreme Court reasoned that the period leading up to publication of a work – while it is being edited and groomed for publication – is a crucial time for authors, so protections for works must be maintained at their highest level and cannot be decreased simply because an author has demonstrated an intent to imminently publish a work.²¹ The court cited the author’s interest in creative control, the property interest that is itself valuable, but also serves to support marketing and publicity to conclude that an “author’s right to control the first public appearance of his undissemminated expression will outweigh a claim of fair use.”²² SSRN is a venue for posting prepublication works that are still undergoing editing. Not only are works subject to change between the time they are posted and the time they are actually published in their intended journals, but journals frequently impose restrictions on publishing works outside the context of SSRN and the author’s own website. Moreover, disseminating an inaccurate summary of an article can be more damaging than circulating a truthful version as it may undermine readers’ confidence or interest in the scholarship and lead them to conclude that the work is not of interest to them, or otherwise not suitable to their purposes. One might assume that under the right circumstances an inaccurately summarized work might also cause offense and reputational damage.

²⁰ *Harper & Row v. Nation Enterprises*, 471 U.S. 539 (1985).

²¹ *Id.* at 555.

²² *Id.*

- *The chatbot violated expectations of those who submit queries to chatbots.* Users of search engines do not expect their queries to be broadcast as news items to other users. It seems reasonable that users of chatbots would have similar expectations for privacy of their interactions with the chatbot unless they are prominently notified that their queries will be published to other users.

All of these experiences taken as a whole suggest that GAI systems are being released in the marketplace without adequate consultation and transparency not just on copyright issues, but on related issues such as privacy and data security.

Question 3: Please identify any papers or studies that you believe are relevant to this Notice. . .
The Office requests that commenters provide a hyperlink to the identified papers.

Although the Copyright Office is to be commended for asking such thought provoking and comprehensive questions, some independent artists worry that their views are being drowned out by more powerful players or lost due to the complexity and overlapping nature of the many inquiries. At a recent roundtable on IP and Social Justice issues one artist commented:

“The big bully on the block, AI has consumed all of the oxygen in the air, and maybe rightfully so. Our work has been scraped and used to create tools that will replace us, without remuneration, authorization, or the barest hint of transparency. Congress calls big tech to tell them what to do about it. How will that work out? USCO calls for comments, but their questions are so technical that they seem to miss the simple basic point. Congress has had some pretty enlightened hearings even. But then what?”²³

This comment reveals the importance of engaging the independent artist community as fully as possible. Modeling questions the Office asks AI proponents to reflect the insights learned from parties who develop community toolkits to assess AI technology for algorithmic equity could be a useful component in shaping the USCO’s further engagement in this area. The toolkit referenced

²³ Benjamin Bierman, *The Other 99 Percent*, Conference Paper, MOSAIC conference on IP and Social Justice (Oct. 27-28, 2023), paper on file with the author.

below is not focused specifically on GAI tools or on copyright – it helps community activists understand, interrogate, assess and pose questions about the intent, effectiveness, impact and oversight of AI systems generally. Adapting the flowcharts and questions posed by the toolkit or using them as a reference point may be helpful to the Office both in ensuring that the Office itself assesses complete project parameters for AI systems it encounters, and also to pose questions that allow the broader community of artists affected by AI to offer useful insights.

- P.M. Krafft, et al., *An Action-Oriented AI Policy Toolkit for Technology Audits by Community Advocates and Activists*, Conference on Fairness, Accountability, and Transparency (March 3-10, 2021), <http://people.csail.mit.edu/pkrafft/papers/critplat-aekit-facct-2021.pdf>

Tom Wheeler, *The Three Challenges of AI Regulation*, Brookings Institution (June 15, 2023), <https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/> provides insightful commentary and some suggestions regarding how AI systems could be better regulated. The suggestions analyze historic calls by Sam Altman, CEO of OpenAI, Brad Smith, President of Microsoft, and Sundar Pinchai, CEO of Google to regulate the activities of the AI industry. The article likewise surfaces the challenges inherent in AI oversight which include: dealing with the velocity of AI developments, understanding what to regulate, and determining who the correct regulatory entities are and how they should act.²⁴

One of the key take-aways advanced by former FCC Chairman Wheeler in the paper is that longneglected harms like violations of personal privacy, manipulation of individual behavior, dissemination of hate, lies and misinformation all can be exacerbated via AI.²⁵ Chairman Wheeler writes “the gateway drug” for digital exploitation (the collection of individual’s personal information), with the large and expanding practices of “siphoning vast amounts of what users of online services have written, videoed, or uttered. The privacy- invading practices that platform companies such as Alphabet/Google, Meta/Facebook, Microsoft, and others have been allowed to pursue have created today’s well recognized problems. Now, as those same companies venture

²⁴ Tom Wheeler, *The Three Challenges of AI Regulation*, Brookings Institution (June 15, 2023), <https://www.brookings.edu/articles/the-three-challenges-of-ai-regulation/>

²⁵ Id.

into AI, these unregulated practices form the basis for further privacy intrusions, including AI-enabled video and audio surveillance of each of us.”²⁶

Finally, *Artists Allege Meta’s AI Data Deletion Request Process is a Fake PR Stunt* <https://www.wired.com/story/meta-artificial-intelligence-data-deletion/> describes the frustrations artists face when trying to use Meta’s new “opt out” tool to request that their data be deleted from use with the latest version of image to text generator Dall-e. Such a system is insufficient to meet a GAI’s obligations under copyright law in any event, because copyright law requires the affirmative consent (opt in) of a copyright owner before a party can exercise any of the copyright owner’s exclusive rights. Moreover, decades of experience with the DMCA’s section 512 safe harbor approach to notice and takedown of works has shown that companies receiving notices to remove infringing works in that context are often disinclined to expeditiously cooperate to remove infringing works from their platforms and that infringing works rapidly repopulate on websites and platforms as soon as they are removed. The article reveals that platform disregard of even self-imposed standards is poised to be even worse in the AI context.

Instead of providing a simple, streamlined means for artists to request removal of their materials from AI training data, artists report that Meta’s announced removal tool is an ineffective “pr stunt” that generates an automated letter stating that their request cannot be processed unless the author can provide evidence of the training data the Meta tool used and the responses from the tool containing the artist’s personal data. Since Meta does not reveal how its AI tool is trained and artists can not possibly be expected to investigate every possible prompt that generates a response incorporating their works or personal data, Wired explains that the tool puts artists in an impossible bind. Further, the article reports that Meta’s spokesperson claimed that the company has no intention of honoring any requests to remove information collected from Meta’s own products and services being used to train its AI models.²⁷ This article illustrates how the problems currently

²⁶ Id.

²⁷ Kate Knibbs, *Artists Allege Meta’s AI Data Deletion Request Process is a Fake PR Stunt* (Oct. 26, 2023) <https://www.wired.com/story/meta-artificial-intelligence-data-deletion/>

extant with Web 2.0 technology are being replicated by powerful players deploying GAI technology.

Question 5: Is new legislation warranted to address copyright or related issues with generative AI? If so, what should it entail? Specific proposals and legislative text are not necessary, but the Office welcomes any proposals or text for review.

As of this morning, the press, public and members of the independent art community are anticipating an announcement from the Biden Administration regarding an Executive Order on AI. Even if the announcement does not specifically mention the work of the Copyright Office, at minimum the Office should conclude this NOI with a comprehensive study summarizing its findings and conclusions. Such a study should consider the impact of proposed legislation related to GAI systems like the No FAKES Act²⁸ and urge Congress to favorably consider it. That Act is not meant to be sufficient to provide a complete solution to the challenges surrounding GAI many commentators have identified, nor does it clarify the legal and regulatory status of existing and developing GAI tools to help authors who might wish to use them in their creative practices.

The Office should therefore consider providing guidance on appropriate copyright and related rights standards Generative AI technology developers should adopt. This may include providing input and suggesting standards for certification to be incorporated into regulatory efforts undertaken by the Biden Administration or further requirements imposed by Congress. As already noted, numerous significant participants in the AI industry have already publicly called for government regulation of their work²⁹, and the Office is aware (and involved in) much work ongoing in sister agencies the USCO frequently collaborates with.³⁰ Because GAI is a general purpose technology that is likely to use the same databases of training materials to train tools with

²⁸ Chris Coons, Marsha Blackburn, Amy Klobuchar, and Thom Tillis, *Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES) Act*, Senator Chris Coons (2023).

²⁹ Courtney Rozen, *AI Leaders Are Calling for More Regulation of the Tech. Here's What That May Mean in the US*, The Washington Post (May 31, 2023), https://www.washingtonpost.com/business/2023/05/31/regulate-ai-here-swhat-that-might-mean-in-the-us/770b9208-ffd0-11ed-9eb0-6c94dcb16fcf_story.html.

³⁰ Kathi Vial, *With Artificial Intelligence Speeding the Innovation Process, What Does That Mean for Invention and a Properly Balanced Patent System?* United States Patent and Trademark Office (April 18, 2023), <https://www.uspto.gov/blog/director/entry/with-artificial-intelligence-speeding-the>.

a wide range of uses, copyright concerns may be implicated by a variety of applications not specifically aimed at the creative sector. Consequently, the Office should not limit its interests only to GAI tools that are directed towards the creative sector, but rather “follow the work” to ensure that any GAI tool trained on copyrighted material meets appropriate copyright standards.

I join the Copyright Alliance comments and thank you for this opportunity to provide additional comments to this notice of inquiry.

Respectfully,

/S/

Sandra Aistars³¹

Clinical Professor

Founder and Director

Mason Arts & Entertainment Advocacy Clinic

George Mason University, Scalia Law School

Sr. Fellow and Sr. Scholar Copyright Policy & Research

Center for Intellectual Property x Innovation Policy

³¹ Affiliations provided for identification purposes only. The author thanks Sofia Lee Frasz, George Mason University, Scalia Law School Class of 2025 for her research assistance.