October 30, 2023

The Honorable Shira L. Perlmutter
Register of Copyrights and
Director of the US Copyright Office
US Library of Congress
101 Independence Ave SE
Washington, DC 20540

Re: ***Notice of Inquiry and Request for Comment re Artificial Intelligence and Copyright*** **[Docket No. 2023-06]**

Dear Register Perlmutter,

BSA | The Software Alliance ("BSA") welcomes this opportunity to provide comments to the United States Copyright Office ("US Copyright Office" or "USCO") in response to the Notice of Inquiry and Request for Comments ("RFC") on Artificial Intelligence ("AI") and Copyright.[1] BSA is the leading advocate for the global software industry before governments and in the international marketplace.[2] BSA has long supported strong and effective copyright protection.

AI can benefit all industry sectors – including content creators – consistent with copyright law. US copyright law is guided by a number of doctrines that have, to date, proven sufficiently flexible to adapt to evolving technologies.

Of course, given the critical importance of AI to long-term US economic growth and national security, it is important for the US Copyright Office to continue monitoring legal developments to ensure that these doctrines continue to foster the development of AI. As stated by Secretaries Raimondo and Blinken, "AI holds an exhilarating potential to improve people's lives and help solve some of the world's biggest challenges, from curing cancer to mitigating the effects of climate

[1] 88 Fed. Reg. 59942 (Aug. 30, 2023) [hereinafter "RFC"].

[2] BSA's members include: Adobe, Alteryx, Asana, Atlassian, Autodesk, Bentley Systems, Box, Cisco, CNC/Mastercam, Databricks, DocuSign, Dropbox, Elastic, Graphisoft, IBM, Informatica, Juniper Networks, Kyndryl, MathWorks, Microsoft, Okta, Oracle, Palo Alto Networks, Prokon, PTC, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Splunk, Trend Micro, Trimble Solutions Corporation, TriNet, Twilio, Unity Technologies, Inc., Workday, Zendesk, and Zoom Video Communications, Inc.

change to solving global food insecurity."[3] As Secretary Blinken also said, "[w]e want American to maintain our scientific and technological edge [in artificial intelligence] because it's critical to us thriving in the 21st century economy."[4]

This critical advancement in technology does not need to be at the expense of artists and rightsholders. In fact, there are many areas, discussed below, in which further discussions may yield more protections for artists and rightsowners.

AI encompasses a broad set of technologies and use cases. Generally speaking, AI systems learn from computational analysis of large amounts of data that can be used to train a model. The model will then make predictions when presented with a query that includes new data. Some generative AI models can take this a step further and generate new data, such as text or a picture or even new code. Regardless of what form of AI is being used, respect for copyrighted works – as well as personal data privacy and trade secret information – should be a cornerstone of responsible AI development.

We provide short answers to the US Copyright Office's four categories of questions below.

First, as regards training AI models, computational analysis typically involves turning data into tokens that are looking for statistical correlations with other tokenized data. Some of that data may be part of a copyrighted work, but the use normally has nothing to do with the expressive content of a work. A book may be used to learn language skills, which are then used for improved database management practices. These uses fall squarely in fair use under copyright law. Furthermore, we understand that there have been voluntary, industry conversations around automated tools to indicate that the rights-owner does not want a website used for training purposes, similar to the current "do not crawl" tools. We encourage further conversations to determine whether a consensus standard is possible.

Second, as regards the copyrightability of works generated using AI systems, the analytical touchstone should be whether human creativity was responsible for the work, regardless of what instrument or technology was used to aid its expression. This is consistent with established practice where software tools are used by artists to create works that are regularly deemed copyrightable, even though their creation was aided by technology. Generative AI should bolster creativity, just as other software applications have long been an important tool of artists and storytellers. Generative AI is used, for example, in word processing for authors and photo enhancements for visual artists; it is used to create special effects in audio-visual works and arranging music for sound recordings. When generative AI is used to enhance human creativity, the resulting work should be protected by copyright.

Third, as regards the risk that AI systems will be used to create infringing works, the copyright laws are sufficiently flexible to determine whether copyright infringement liability exists. Stated differently, whether the output is infringing is typically not related to the technology used. Where an output is infringing, there should clearly be liability regardless of whether AI is used in

[3] Gina Raimondo and Anthony Blinken, *To Shape the Future of AI, We Must Act Quickly*, Financial Times (July 24, 2023), at: https://www.ft.com/content/eea999db-3441-45e1-a567-19dfa958dc8f

[4] Speech by Secretary of State Anthony Blinken, A Foreign Policy for the American People (March 3, 2021), at: https://www.state.gov/a-foreign-policy-for-the-american-people/; *See also,* Speech by National Security Advisor Jake Sullivan (Sept. 16, 2022), at: https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/09/16/remarks-by-national-security-advisor-jake-sullivan-at-the-special-competitive-studies-project-global-emerging-technologies-summit/ (noting that, "leadership in … computing-related technologies… is a national security imperative.")

producing the infringing copy. This is also consistent with established law which focuses on whether a work infringes and is less concerned with the particular method used to create the work.

Fourth, while copyright law can address infringement, we agree that it is worth further exploring whether a federal law protecting against unauthorized digital replication of name, image, likeness, or voice is warranted. While there are state laws that provide some related protection, we encourage the Copyright Office to evaluate whether a harmonized, federal law would better protect artists.

I. Discussion

Questions 6-14 ask about how data is used to train AI systems. As discussed in more detail below, there are different approaches to AI training, but this is the core of each: turning bits of data into tokens and finding statistical correlations from the tokens. The bits of data may be part of a licensed data set, they may be freely available online, or they may be part of a combination of different sets.

Training data may or may not be part of a copyrighted work, since there are no formalities to copyright protection and all expressive works are protected. The training set, however, is used to enable studying of non-expressive data it contains. This enables the system to learn, for instance, how language functions or about spatial relationships. While this form of data analysis qualifies as a fair use under the Copyright Act, there may be situations in which the copyright owner or publisher wants to prevent the data use for training purposes. There are ongoing industry conversations about how to create metatags to create a signal on an otherwise freely available website, and the Office should encourage those conversations.

A. Artificial Intelligence in a Copyright Context

AI machine learning encompasses a vast array of technologies developed or deployed for use in a variety of different industries and applications. Machine learning[5] depends upon computational analysis of training data[6] to identify correlations, patterns, or other metadata in order to develop a model that can make predictions or recommendations based on future data inputs.[7] More

[5] The USCO RFC defines machine learning as "[a] technique for building AI systems that is characterized by the ability to automatically learn and improve on the basis of data or experience, without relying on explicitly programmed rules. Machine learning involves ingesting and analyzing materials such as quantitative data or text and obtain inferences about qualities of those materials and using those inferences to accomplish a specific task. These inferences are represented within an AI model's weights."

[6] The USCO RFC defines training datasets as, "[a] collection of training material (as defined below) that is compiled and curated for use in machine learning. Examples of training datasets include BookCorpus, ImageNet, and LAION." The USCO further defines training material as, "individual units of material that are used for purposes of training an AI model. They may include a combination of text, images, audio, or other categories of expressive material, as well as annotations describing the material. An example of training material would be an individual image and an associated text ''label'' that describes the image."

[7] *See* BSA | The Software Alliance, *Confronting Bias: BSA's Framework to Build Trust in AI* (2022), at: https://ai.bsa.org/confronting-bias-bsas-framework-to-build-trust-in-ai; *See also,* National Institute of Standards and Technology, *NIST Risk Management Framework* (Jan. 2023), at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (describing an AI system "as an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments.")

recently, generative AI has emerged, allowing for the creation of new outputs in textual, visual, or aural formats.[8]

1. Applications of AI

We offer below a few (widely recognized) examples of insights, predictions, and other outputs derived from computational analysis in the machine learning context:

- Automated flight management and air traffic control based on computational analysis of meteorological conditions, real-time fuel consumption, aircraft operational data, nearby air traffic conditions, airport congestion, and numerous other data elements.[9]

- Identification of chemical and cellular anomalies for early diagnosis, prevention, and treatment in the fields of oncology, autoimmune disorders, and Parkinsons and Alzheimers disease.[10]

- By integrating generative AI into security operations, organizations can effectively identify and address security anomalies, as well as detect and mitigate potential threats.

- Predictive climate modeling based on computational analysis of satellite data, weather station data, topographical information, and various IoT and sensor data.[11]

---

[8] The USCO RFC defines generative AI as "[a]n application of AI used to generate outputs in the form of expressive material such as text, images, audio, or video. Generative AI systems may take commands or instructions from a human user, which are sometimes called ''prompts.'' Examples of generative AI systems include Midjourney, OpenAI's ChatGPT, and Google's Bard."

[9] *See e.g.,* M. Durgut, *Artificial Intelligence and Air Traffic Control*, Aviationfile.com website (Jan. 2023), at: https://www.aviationfile.com/artificial-intelligence-and-air-traffic-control/#:~:text=One%20of%20the%20primary%20applications%20is%20to%20help,make%20informed%20decisions%20on%20routing%20and%20scheduling%20flights; Degas et al., *A Survey on Artificial Intelligence and explainable AI in Air Traffic Management*, 12 Applied Sciences 1295 (2022), at: https://www.mdpi.com/2076-3417/12/3/1295; Hanneke Weitering, *How Artificial Intelligence is Transforming Aviation*, Futureflight.aero website (2023), https://www.futureflight.aero/news-article/2023-07-13/beyond-automation-how-artificial-intelligence-transforming-aviation

[10] *See e.g.,* Hunter et al., *The Role of Artificial Intelligence in Early Cancer Diagnosis*, 14(6) Cancers 1524 (2022), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8946688/; Stafford et al., *A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases*, 3 NPJ - Digital Medicine 30 (2020), at: https://www.nature.com/articles/s41746-020-0229-3; Diogo et al., *Early diagnosis of Alzheimer's disease using machine learning*, 14 Alzheimers Research and Theory 107 (2022), at: https://alzres.biomedcentral.com/articles/10.1186/s13195-022-01047-y; Ahsan et al., *Machine-Learning-Based Disease Diagnosis: A Comprehensive Review*, 10(3) Healthcare Basel 541 (2022), at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8950225/; Luchini et at., *Artificial Intelligence in Oncology*, 126 British J. of Cancer 1 (2022), at: https://www.nature.com/articles/s41416-021-01633-1; Yumar et al., *Artificial intelligence in Disease Diagnosis*, 14(7) J. Ambient Intell Humaniz Comput. 8459 (2023), at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8754556/

[11] Schneider et al., *Harnessing AI and computing to advance climate modelling and prediction*, 13 Nature Climate Change 887 (2023), at: https://www.nature.com/articles/s41558-023-01769-3; World Economic Forum, *The role of machine learning in helping to save the planet* (2021), at: https://www.weforum.org/agenda/2021/08/how-is-machine-learning-helping-us-to-create-more-sophisticated-climate-change-models/; Kaak et al., *Aligning artificial intelligence with climate change mitigation*, 12 Nature Climate Change 518 (2022), at https://www.nature.com/articles/s41558-022-01377-7; Xin et al., *Artificial Intelligence for Climate Change Risk Prediction, Adaptation, & Mitigation*, Ecological Processes (2021), at: https://www.springeropen.com/collections/AICC; Chantry et al., *Opportunities and challenges for machine learning in weather and climate modelling*, 379 Phil. Trans. R. Soc. 83 (2020), at: https://doi.org/10.1098/rsta.2020.0083 (2020).

- Improved carbon tracking and mitigation based on computational analysis of transportation logs, meter readings, fuel purchase records, atmospheric pollution tracking, and visual monitoring of power plants and other facilities, and other data sources.[12]

- Computational analysis to map vulnerable seaside areas to produce cyclone risk maps and guide investment plans for cyclone shelters, schools, health facilities, and other infrastructure for disaster planning and survivability.[13]

- Predictive typing and other office productivity solutions (e.g., an "auto-complete" function that suggests the letters "…cerely yours" after the typist inputs the letters, "sin"),[14] or the creation of sound effects or special effects to assist creators and film producers in developing new artworks.[15]

While the use cases are diverse, the elements of training each are very similar and discussed below.

2. The AI Development Life Cycle

The AI development life cycle typically includes the following steps:

- Project Conception: Often, AI models are designed as general-purpose tools, such as a natural language processing model, that may later be customized for multiple particular tasks. Sometimes, an AI model is developed with a specific task in mind. In that situation, the AI development team must formulate the "problem" that a system is intended to address and map the structure and target variables that the system is intended to predict. For instance, a fitness app that analyzes a consumer's heart rate to monitor for irregularities that might predict whether that person is at risk of a stroke or heart disease (i.e., the target variable).

- Raw Data Identification: Often, AI models are trained on massive amounts of unlabeled data through self-supervised learning to serve as general purpose tools, such as a natural language processing model, that may later be customized for multiple particular tasks. In other circumstances where the model is being developed for a specific task, the AI development team must identify a relevant universe of "raw data" that will subsequently be transformed and structured. Data sources are as diverse as the potential applications of machine learning AI and may include everything from machine-to-machine data (e.g., satellite transmission data) and international trade statistics to published materials, blog

[12] *See e.g.,* Global Data Alliance, *Cross-Border Data Transfers & Environmental Sustainability* (2023) (internal citations omitted), at: https://globaldataalliance.org/wp-content/uploads/2023/04/04192023gdacbdtsustainability.pdf

[13] *See id.*

[14] *See e.g.,* S. Ashraf, *Desmystifying Autocomplete,* Towards Data Science (2020), at: https://towardsdatascience.com/index-48563e4c1572; A. Wickramarachchi, *Machine Learning: Word Embedding and Predicting,* Towards Data Science (2020), at: https://towardsdatascience.com/machine-learning-word-embedding-and-predicting-e603254e4d7b

[15] *See e.g.,* D. Nelson, *AI Researchers Design Program To Generate Sound Effects For Movies and Other Media*, Unite AI Website (2022), https://www.unite.ai/ai-researchers-design-program-to-generate-sound-effects-for-movies-and-other-media/

posts, website comments, and chat room logs. "Raw data" is frequently "messy," requiring significant work to transform the data into a usable form, as outlined below.

- Preparing the Data Set: The AI development team must modify the "raw data" so that it can be semantically understood by the machine and used to train the model. During this process, the team will revise, clean, and normalize the data as necessary. Data is typically semantically and structurally transformed through "tokenization," which involves breaking down a piece of text or data into smaller units (or "tokens") for purposes of computational analysis. Tokenization is necessary to improve the reliability, quality, and suitability of the data for analysis, helping to address quality challenges, such as missing values, duplicates, outliers, and inconsistent formatting across the entire data set. Training data will often ultimately appear as a sequence of tokens that stretches across the entire AI corpus.

- Model definition: After input data has been suitably processed, the AI development team must establish the system's underlying architecture. This includes identifying the variables (i.e., "features") in the training data that the algorithm will evaluate as it looks for patterns and relationships as the basis of a rule for making future predictions. It also includes selecting the type of algorithmic model that will power the system (e.g., linear regression, logistic regression, deep neural network.)[16] Once the data is ready and the algorithm is selected, the team will train the system to produce a functional model that can make predictions about future data inputs.

- Model Validation, Testing, and Revision. After the model has been trained, it must be validated to determine if it is operating as intended and tested to demonstrate that the system's outputs fall within expected parameters and do not contain unexpected errors or unintended bias. Based on outcome of validation and testing, the model may need to be revised and refined to mitigate these risks.

B. The Use of Copyrighted Works to Train AI Models

As discussed above, computational analysis is typically applied to a large training data corpus that may comprise millions or billions of tokenized data elements. Depending on how the model is trained, data accessible over the Internet may be collected as part of the raw data set that is transformed into that corpus.

This raw data may include copyrighted works because a substantial portion of the internet is potentially subject to copyright protection, which has a low threshold to establish "originality" and which provides that copyright arises automatically upon a work's creation, even if it the work is

[16] *See* BSA AI Bias Framework, *infra* note 7.

not registered.[17] Importantly, however, not all of the material online is subject to copyright, in part because copyright protection does not extend to facts, ideas, or mathematical concepts.[18]

Computational analysis may involve two sets of reproductions that potentially implicate the Copyright Act: (1) reproductions necessary to create a corpus of "training data," and (2) transient reproductions that are incidental to the computational process of training the AI model. In each case, the reproductions are "intermediate" in the sense that they are not visible or otherwise made available to the public. Instead, the reproductions are the necessary byproduct of a technical process that is aimed at identifying non-copyrightable information *about* the underlying corpus of works – i.e., the correlations and patterns that inform the creation of the AI model and enable it to make predictions based on future data inputs. Such intermediate, non-expressive reproductions do not compete with and have no impact on the economic interests that copyright is intended to protect.

Furthermore, computational analysis does not involve the consumption of any copyrighted works for their expressive content. Rather, such analysis involves mathematical calculations of probabilities, correlations, trends, and other patterns across the entire tokenized data set. Such analysis seeks to understand only the mathematical patterns (e.g., the relationships of specific tokens in relation to other tokens) distributed across the entire data set. These mathematical patterns are themselves not expressive content protected by copyright law.

As noted below, established legal doctrine ensures that such reproductions are generally permissible.

1. Computational Analysis of AI Training Data Is a Fair Use

The Constitution authorizes Congress to create a framework for protecting intellectual property in order to "promote the progress of science and useful arts."[19] The fair use doctrine is a critical safety valve that ensures the Copyright Act remains consistent with its Constitutional purpose.[20] Codified in Section 107 of the Copyright Act, the fair use doctrine allows for unauthorized uses of copyrighted works in certain circumstances, providing courts with a flexible mechanism for balancing "the interest of authors and inventors in the control and exploitation of their writings and

[17] *See e.g.,* 17 USC § 102(a) ("Copyright subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.")

[18] *See* Baker v. Selden, 101 U.S. 99 (1879); *see also* "Article 9.2 of the TRIPS Agreement provides that copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such. In other words, copyright protection does not cover any information or ideas contained in a work; it only protects the original way in which such information or ideas have been expressed. Thus everyone is free to use the information contained in a work, including for the purpose of creating new works. For example, the idea behind a detective novel is not protected as such, but an unauthorized reproduction of that novel that is an expression of the idea is prohibited. This principle, commonly referred to as the 'idea/expression dichotomy', has always been present in copyright doctrine, although it had not been explicitly set out in the provisions of the Berne Convention. Article 9.2 of the TRIPS Agreement is therefore the first explicit confirmation of the principle in multilateral IP law." WTO, Overview of the Agreement on Trade-Related Intellectual Property Rights (2023), at: https://www.wto.org/english/tratop_e/trips_e/ta_docs_e/modules2_e.pdf

[19] U.S. Const. Art I, § 8, Cl. 8.

[20] *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 575 (1994).

discoveries on the one hand, and society's competing interest in the free flow of ideas, information, and commerce on the other hand."[21]

It is impossible, and unnecessary in light of fair use law's focus on facts of a specific case, to draw a generalized conclusion that *all* applications of AI involving the reproduction of copyrighted works will qualify as fair uses. But the case law suggests strongly that the fair use doctrine will accommodate the creation and use of AI training databases for the purposes of analyzing a large collection of individual works to identify patterns, correlations, and other metadata to develop an AI model that makes predictions about future data inputs.

By way of illustration, this means that an AI developer seeking to create a natural language processing model – such as an AI-driven predictive typing model – can rely on publicly available textual material (ranging from the letters of Flannery O'Connor to anonymous website commentary) to create the training database. In such a scenario, the AI developer would not be reproducing this text for its expressive purpose. Rather, the reproductions would be made solely for the purpose of extracting unprotected information *about* the English language – i.e., the correlations, patterns, and relationships among the 26 letters of the alphabet and the 1 million English language words, as they appear in thousands of stock phrases, figures of speech, similes, metaphors, grammar patterns, and common linguistic formulations and expressions. Neither these letters, words, and phrases, nor the mathematical patterns among them across thousands or millions of writings, are copyright protectable subject matter.

Furthermore, the ultimate use of the computational analysis applied to the data set is a transformative use of the original content. Auto-complete functionality in predictive typing software comprises a new creation in the form of software code that is transformatively distinct both from the entire AI training corpus and from any single work within that corpus. Software code that can suggest the endings of commonly used phrases is many steps removed from any copyrighted works found in the underlying raw data. Finally, such functionality simply does not compete with any copyrighted works in any manner that copyright is intended to protect. Accordingly, the AI developer can safely rely on the fair use doctrine to construct an AI training database using photographs to which she has lawful access.

In addition to the reproductions that may be needed to create an AI training database, reproductions may also be made when the training data undergoes the computational analysis that occurs during the machine learning process. These fleeting reproductions are an inevitable byproduct of working with digital media, which must be loaded into the random access memory (RAM) of a computing device to be accessed, analyzed, manipulated or even deleted.[22] In some circumstances, the creation of RAM copies may constitute a reproduction that is cognizable under the Copyright Act. However, the fleeting nature of the RAM copies involved in the machine learning process are unlikely to meet the Copyright Act's threshold for "fixation" and are therefore unlikely to be considered "copies" that implicate the reproduction right.[23]

---

[21] Harper & Row Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 580 (1985) (quoting Sony Corp. of Am. V. Universal City Studios, Inc., 464 U.S. 417, 429 (1984)).

[22] *See* Aaron Perzanowski, *Fixing RAM Copies*, 104 NW. U. L. REV. 1067 (2010); Jule L. Sigal, *Copyright Infringement Was Never This Easy: RAM Copes and Their Impact on the Scope of Copyright Protection for Computer Programs*, 45 CATH. U. L. REV. 181 (1995).

[23] The reproduction right affords copyright owners the exclusive right to make "copies" of their works. 17 USC § 106. For a copy to trigger the reproduction right, it must be "fixed" in a manner that allows the work to be "perceived,

The Court's recent decision in *Andy Warhol Foundation v. Goldsmith* does not change the analysis. In *Warhol*, the Court looked at the specific use alleged to infringe and assessed the justification for the use to determine whether it supported a finding of fair use. The Court essentially held that fair use does not protect a licensee who has a license to use a work once, but uses the work additional times and then licenses those subsequent works. As described above, in the case of training an AI system, the use of data in any underlying copyrighted work is for training purposes and the trained model neither includes a copy of the underlying work nor is it a competing work.

2. Voluntary Measures and Consensus Solutions

Finally, as outlined in RFC question 9, there have been calls to create mechanisms that restrict access to works online for any purpose, including training an AI system. It should be remembered that automated "do not crawl" tools can prevent access to works online for any purpose, including training an AI system. We support further voluntary conversations between creators and AI developers and deployers to arrive at effective, consensus technical mechanisms.

C. The Copyrightability of Material Generated by an AI System

Questions 18-21 ask about copyrightability. For both policy and legal reasons, copyright protection should be available for works generated by a human using AI. As long as human creativity was responsible for the work, the tool used in aiding the expression should not matter.

Inquiries regarding copyrightability will turn on a close examination of the degree of human creativity. Artificial intelligence is increasingly being used as a tool in all categories of creative works. A decision to limit copyrightability when AI is used would significantly chill adoption of AI solutions. In turn, this would limit creative expression, contrary to Constitutional imperative, and impact US leadership on AI, contrary to the Administration's objectives.

Works that emerge as outputs of AI systems and meet the human creativity requirement should continue to be eligible for copyright protection. In most cases, AI systems will function as tools used by human authors to execute upon their creative vision. For instance, photographers will use AI-enabled tools to automate the tedious process of editing their images,[24] architects will use AI to augment their designs to enhance their energy efficiency,[25] and filmmakers will use AI to ensure that the movement of their animated characters appear more life-like.[26] In each of these cases the creative contribution of the human user makes it easy to conclude that the output would be copyrightable.

The use of generative AI should not change the analysis. Certainly, there will be cases at either extreme, in which it is either clear that there is no spark of human creativity involved or, on the other hand, generative AI was not part of the creative expression. We encourage the Copyright

---

reproduced, or otherwise communicated." 17 USC § 101. A copy is considered "fixed" when its "embodiment" is "sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for more than a period of transitory duration." *Id.* A copy therefore implicates the reproduction right only when it is (1) embodied in a medium and (2) the embodiment subsists for more than a "transitory" duration.

[24] https://theblog.adobe.com/adobes-general-counsel-makes-the-case-for-ai/.

[25] https://www.autodesk.com/redshift/machine-learning-in-architecture/.

[26] https://theblog.adobe.com/state-of-ai-in-animation/.

Office, in the first instance, to take a broad view of copyrightability in the registration process. The Copyright Act provides courts with sufficient flexibility to closely examine these considerations when needed.

D. Potential Liability for Infringing Works Generated Using AI Systems

Questions 22-27 relate to potentially infringing works generated using AI systems. BSA offers the following observations.

1. Current Copyright Law Protects Copyright Holders from Infringement, Including in Cases Arising from AI Generated Content

Copyright holders should have full and effective remedies when their rights are infringed. This principle applies equally to outputs generated using AI systems and outputs generated in other ways. Whenever such infringement is found, it is critical to fully compensate artists and creators for any damages caused.

In our view, existing copyright law should prove adequate to address questions of infringement. In most AI use cases, such as those described above, the output of an AI system will not implicate copyright at all. However, AI, like other technologies, could be used to create infringing material. Generally speaking, in those situations, infringement liability would be premised upon proof that the allegedly infringing output is based on "actual copying" and is "substantially similar"[27] to the copyrighted work, such that an "ordinary reasonable person would fail to differentiate between the two works" in view of the "qualitative[ly] and quantitative[ly] significant" similarity between the works.[28] In cases in which the copyright claims are "thin," because there is only one way, or there are only a few ways, to represent specific facts, it would be necessary to prove "virtual identity" between the output and copyrighted work, consistent with prevailing legal norms.[29] Many users of generative AI will be small businesses experimenting with new tools, and enforcement decisions should take account of this. Liability would also arise in other appropriate cases, such as those involving derivative works.

Plaintiffs may also seek to bring infringement actions against providers of AI-related services. Here too, existing copyright doctrines should prove adequate in evaluating and apportioning liability. If a plaintiff demonstrates that a direct infringement has occurred, courts will evaluate whether the service provider should be deemed "contributorily" and/or "vicariously" liable for its users' conduct. In the context of contributory liability claims, the staple article of commerce doctrine will ensure that providers of AI services with substantial non-infringing uses are not liable for their users' infringing activity unless there is evidence that the service was made available "with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement."[30]

---

[27] *See generally*, University of Michigan Law Library, Substantial Similarity Guide (2023), at: https://guides.lib.umich.edu/substantial-similarity/overview (accessed Oct. 2, 2023).

[28] *See e.g., Williams v. Kaag Manufacturers*, 338 F.2d 949 (9th Cir. 1964)

[29] *See e.g., Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435 (9th Cir. 1994)

[30] *See e.g., Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.,* 545 US 913 (2005).

2. AI Models Trained on Sufficiently Large Data Sets Are Less likely to Produce Infringing Outputs

While it is important not to conflate training data with the output of an AI system, it is worth noting that the more data available for training, the less likely the system will produce a copy or derivative of an input (in the absence of a user's intent to infringe). In any well-designed AI model trained on a sufficiently large data set, computational analysis should never (or only in the rarest of circumstances) produce outputs that are "substantially similar," let alone "virtually identical," to any specific copyrighted work. As Professor Matthew Sag of Emory University has explained,

> At the moment, memorization [i.e., the accidental replication of a particular work from a training set] is an edge case. For the most part, the link between the training data and the output of generative AI is attenuated by a process of decomposition, abstraction, and remix. Generally, pseudo-expression generated by large language models does not infringe copyright because these models "learn" latent features and associations within the training data, they do not memorize snippets of original expression from individual works.[31]

Former General Counsel of the US Copyright Office Sy Damle has also explained that:

> [A]ttempts to build AI models using smaller sets of licensed or public domain material will lead to models that are less effective and, ironically, more likely to (inadvertently) create outputs that simply regurgitate their training data.[32]

Additionally, as noted above, some AI developers and deployers are already taking steps to engage with artists and creators on how to support their work in a changing digital environment and taking steps to limit misuse of an AI system, such as limiting what prompts can be used.

E. The Treatment of Generative AI Outputs that Imitate the Identity or Style of Human Artists

Questions 30-32 ask about other legal rights that may apply to AI-generated material, including whether Congress should establish a new federal right, similar to state law rights of publicity.

In our view, it is worth exploring how federal legislation could work to afford athletes, musicians, singers, actors, and other artists, performers, and public figures with the opportunity to prevent others from unauthorized exploitation of their name, image, likeness, or voice.

There is a strong legislative policy rationale for the adoption of uniform nationwide standards in this area: First, "right of publicity" protections currently exist only as a patchwork of state statutes and common law doctrine.[33] Second, artists, performers, and public figures have a strong interest

---

[31] *Matthew Sag, Copyright Safety for Generative AI*, 61 Hous. L. Rev. 101 (Aug. 10, 2023), at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4438593

[32] *Statement of Sy Damle re "Artificial Intelligence and Intellectual Property Law - Interoperability of AI and Copyright Law,"* Testimony before the US House of Representatives Committee on the Judiciary, Subcommittee on the Courts, Intellectual Property, and the Internet (May 17, 2023). "Critically, the ability of AI models to duplicate training data is a bug, not a feature." *Id.*

[33] For a survey of the US states and foreign jurisdictions that protect the right of publicity (also referred to as the right of personality), see International Trademark Association, Right of Publicity Committee, Right of Publicity State of the

in maintaining control over their names, images, likenesses, or voices because generative AI and other digitally replicative technologies can imitate those attributes of personal identity without authorization, causing potential commercial harm to the artists, performers, and public figures who depend upon their name, image, likeness, and voice to support themselves. This may address many of the legitimate concerns raised in this area.[34]

In evaluating possible federal legislation, BSA encourages the US Copyright Office to evaluate US state statutes and common law to ensure that any federal legislation is broadly consistent and coherent with existing norms and practice.

II. Conclusion

BSA thanks the US Copyright Office for the opportunity to provide comments in response to the USCO's RFC on AI and Copyright. BSA validates the resilience and adaptability of existing US law in fulfilling its constitutional mandate to "promote the Progress of Science and useful Arts" while securing to creators "exclusive right(s) to their respective writings and discoveries." The United States can maintain the US "leadership in technology and innovation [that] has long underpinned our economic prosperity and military strength,"[35] while also protecting the interest of creators.

As discussed above, the non-consumptive computational analysis of non-copyrightable content in AI training data sets is a "fair use" under any reasonable interpretation of US copyright law. At the same time, BSA supports multi-stakeholder efforts relating to AI training processes as well as efforts to minimize the risk of infringement. And to the extent that infringement occurs, BSA strongly supports fully protecting content creators. Finally, BSA supports exploring how federal legislation could better protect against improper digital replication of a person's name, image, likeness, or voice in a manner that competes with his or her professional and commercial interests.

Thank you for the opportunity to share our views on these important issues.

Sincerely,

Aaron Cooper
Vice President, Global Policy

Law Survey (2019), at: https://www.inta.org/wp-content/uploads/public-files/advocacy/committee-reports/INTA_2019_rop_survey.pdf

[34] *See generally Testimony of Dana Rao re "Artificial Intelligence and Intellectual Property – Part II: Copyright,"* Testimony before the US Senate Committee on the Judiciary, Subcommittee on Intellectual Property (July 12, 2023), p. 6, at https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_rao.pdf.

[35] The White House, *National Security Strategy* (Oct. 2022), at: https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf