BEFORE THE
NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION

AI Accountability Policy Request for Comment	**Docket No. 230407–0093**

The Copyright Alliance appreciates the opportunity to submit the following comments in response to the request for comments[1] (RFC) published by the National Telecommunications and Information Administration (NTIA) in the Federal Register on April 13, 2023, regarding self-regulatory, regulatory, and other measures and policies that are designed to provide assurance that artificial intelligence (AI) systems are legal, effective, ethical, safe, and otherwise trustworthy.

The Copyright Alliance is a non-profit, non-partisan public interest and educational organization representing the copyright interests of over 2 million individual creators and over 15,000 organizations in the United States, across the spectrum of copyright disciplines. The Copyright Alliance is dedicated to advocating policies that promote and preserve the value of copyright, and to protecting the rights of creators and innovators. The individual creators and organizations that we represent rely on copyright law to protect their creativity, efforts, and investments in the creation and distribution of new copyrighted works for the public to enjoy.

The Copyright Alliance supports the responsible, respectful, and ethical development and use of AI technologies and a thriving and robust AI economy. Many of our members are already using or plan to use AI to aid in the creation of a wide range of works that benefit society, and some are themselves developers of AI technologies. We commend the NTIA, the Department of Commerce (DOC), and each federal agency involved in the Administration's coordinated effort to advance trustworthy and responsible development of AI applications. The RFC references the

---

[1] AI Accountability Policy Request for Comment, 88 Fed. Reg. 22433 (issued Apr. 13, 2023).

DOC's focus on "solidifying U.S. leadership in emerging technologies, including AI." As the Department recognizes—through its statement that the "United States seeks to promote the development of innovative and trustworthy AI systems that respect human rights, [and] democratic values, and are designed to enhance privacy protections" and other actions—the U.S. position as a leader in AI is not only about *what* we are able to achieve, but just as importantly *how* we get there.

Respect for copyright law plays a critical role in the trustworthy and responsible development of AI systems. We are pleased that the U.S. Copyright Office and the U.S. Patent and Trademark Office (USPTO), as subject matter experts, are each conducting studies on AI and copyright[2] and intellectual property[3] more broadly. Questions about infringement, copyrightability, and liability related to the ingestion of copyrighted works and the output of works created using AI are topics that the Copyright Office has already begun exploring through its series of listening sessions as well as its recent registration guidance[4] and will continue to explore as its study on copyright and AI progresses. To the extent that copyright issues arise in the course of the NTIA study on AI accountability, we trust that NTIA will defer to the Copyright Office and USPTO's guidance on these issues. We, along with a diverse group of other stakeholders, have been actively involved in the Copyright Office and USPTO studies on AI and will continue to engage with these offices as those studies progress.

While we recognize that there are other important issues related to AI accountability, our comments are limited exclusively to issues directly related to AI and the use of copyrighted works. We appreciate the opportunity to respond to those inquiries raised in the NTIA study that implicate copyright issues.

---

[2] *See* U.S. Copyright Off., *Copyright Office Launches New Artificial Intelligence Initiative*, NEWSNET ISSUE 1004 (Mar. 16, 2023), https://copyright.gov/newsnet/2023/1004.html [https://copyright.gov/newsnet/archive/]; *see also* Symposium, *Copyright in the Age of Artificial Intelligence*, U.S. COPYRIGHT OFF. (Feb. 5, 2020) https://www.copyright.gov/events/artificial-intelligence/?loclr=blogcop.

[3] *See Public Views on Artificial Intelligence and Intellectual Property Policy*, U.S. PAT. & TRADEMARK OFF. (Oct. 2020), https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf

[4] Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 88 Fed. Reg. 16190 (proposed Mar. 16, 2023), https://copyright.gov/ai/ai_policy_guidance.pdf.

Copyrighted Works Are Not "Data"

There are numerous references to and questions about "data" throughout the RFC. As an initial matter, we want to highlight an important distinction between *data* and copyrighted works. In discussing the massive amount and array of material ingested by AI systems, some people have begun to incorrectly lump copyrighted works under the umbrella term "data." We want to make it clear that *copyrighted works are not data.*

This is not just a matter of semantics. Instead, it is an issue that gets to the core of our concerns relating to AI and copyright. The term data refers to (1) "*factual information* (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation;" (2) "*information* in digital form that can be transmitted or processed;" or (3) "*information* output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful" (emphasis added). [5]

By contrast, copyrighted works—books, music, movies, photographs, paintings, sculptures, etc.—are works of expression. In fact, to be protected under copyright law, a work cannot be mere data—i.e., facts or information. To garner copyright protection, a work *must be* an authored work of expression; as the Supreme Court has made clear, "facts are not copyrightable." *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340, 344 (1991); *see also Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 547 (1985) ("[N]o author may copyright facts or ideas. [17 U.S.C.] §102."). To mislabel a copyrighted work as mere "data" is to strip it of the critical essence by which it avails itself of copyright protection: its expressive value and human creativity.

While there are important discussions to be had about the ingestion of *data* by AI systems, those discussions differ in substance, and must remain separate, from discussions about the ingestion of *copyrighted works*. It is therefore vital that the term "data" be reserved for facts and information and not be used to refer to copyrighted works. As an organization whose mission focuses on promoting and preserving the value of copyright, and protecting the rights of creators and innovators, our comments will focus specifically on AI and copyrighted works.

---

[5] *Data*, MERRIAM-WEBSTER, https://perma.cc/7F56-XRGU.

Intellectual Property and the AI Accountability Ecosystem

        Question 27 asks about "the role of intellectual property rights… in fostering or impeding a robust AI accountability ecosystem." An accountable AI ecosystem requires respect for copyright law and the rights of creators and copyright owners. As AI technology continues to evolve and questions arise about how copyright laws apply to the ingestion of copyrighted works by AI, it is critical that the underlying goals and purposes of our copyright system are upheld and that the rights of creators and copyright owners are respected. Long-standing copyright laws and policies must not be cast aside in favor of new laws or policies obligating creators to essentially subsidize AI technologies. There is no "AI exception" to copyright law, nor should policymakers or courts create one. Likewise, when formulating new AI laws and policies, it is essential to respect the rights of creators and copyright owners and whether and how they choose to exercise their rights. Indeed, the National Institute of Standards and Technology (NIST) in its AI Risk Management Framework,[6] as well as the G7 in a [Ministerial Declaration](#) of its Tech Ministers' Meeting[7] acknowledge copyright/intellectual property considerations and the importance of human centricity in the context of developing a trustworthy and accountable AI ecosystem. The AI Risk Management Framework also specifically notes that "training data may also be subject to copyright and should follow applicable intellectual property rights laws" and identifies intellectual property infringement as a possible risk when developing AI.

        Independent-to-large-scale creators and copyright owners produce high-quality works that are often ideal for ingestion by AI systems in order to generate high-quality output. Many creators and rightsholders, particularly publishers and image/media licensors, already license their copyrighted works for commercial AI uses and many of those that do not are on the cusp of doing so. Copyright law incentivizes those creators and rightsholders to lawfully enhance and aggregate their copyrighted works for that purpose—such as through semantic enrichment, metadata tagging, content normalization, and data cleanup. Where a copyright owner offers licenses for the purpose of ingestion, it is essential that these licenses be respected by any

---

[6] Artificial Intelligence Risk Management Framework (AI RMF 1.0), NAT'L INST. STANDARDS & TECH. (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[7] Group of Seven [G7], Ministerial Declaration of 30 April 2023, https://g7digital-tech-2023.go.jp/topics/pdf/pdf_20230430/ministerial_declaration_dtmm.pdf.

copyright or AI legal regime.[8] The marketplace should continue to properly value and incentivize creativity, and AI policy should not interfere with the right of copyright owners to license, or choose not to license, their works for AI uses.

While some AI developers may attempt to downplay or even deny the role of copyright in the AI accountability ecosystem by suggesting that AI systems are copying only *facts* and *information*, or that the ingestion of copyrighted works by AI categorically constitutes fair use, these assertions are not correct. Ingestion of copyrighted works by AI systems without authorization is a major concern of copyright owners big and small and for all types of creative works, and, unless it is excused by fair use (which can only be definitively determined by a court), constitutes copyright infringement on a massive scale. In some cases, in addition to infringements that occur during the ingestion phase, works generated by AI may also infringe works ingested by the AI system.

Some AI developers assert that AI innovation will be impeded if the ingestion process is not deemed to be categorically fair use. This false narrative purposely omits the fact that in many cases licenses are available for ingestion. As history has shown us, creators and copyright owners are usually willing to license their works if the parties can agree on appropriate terms and compensation; that is, of course, how creators typically earn a living. Copyrighted works provide immense value to AI developers, and they can and should pay for that value—as many today are already doing. In other words, when properly applied, copyright law sets the conditions for the market to prevail.

Transparency Regarding the Ingestion of Copyrighted Works

Questions 15, 19, 20, and 21 touch on topics related to transparency. Transparency and accountability/trustworthiness are two sides of the same coin, and adequate transparency regarding ingestion of copyrighted works goes a long way in helping to ensure that copyright

---

[8] The existence of a licensing market weighs against a finding that copying without the permission of the copyright owner is excused by the fair use defense. *See Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 929 (2d Cir. 1994) ("[I] is indisputable that, as a general matter, a copyright holder is entitled to demand a royalty for licensing others to use its copyrighted work, see 17 U.S.C. § 106 (copyright owner has exclusive right "to authorize" certain uses), and that the impact on potential licensing revenues is a proper subject for consideration in assessing the fourth [fair use] factor….").

owners' rights are respected. Best practices from corporations, research institutions, governments, and other organizations that encourage transparency around AI ingestion already exist that enable users of AI systems or those affected by its outputs to know the provenance of those outputs.[9] In particular, except where the AI developer is also the copyright owner of the works being ingested by the AI system,[10] it is vital that AI developers maintain records of which copyrighted works are being ingested and how those works are being used, and make those records publicly accessible as appropriate (and subject to whatever reasonable confidentiality provisions the parties to a license may negotiate).[11]

Adequate and appropriate transparency and record-keeping benefit both copyright owners and AI developers in resolving questions regarding infringement, fair use, and compliance with licensing terms. Those practices can also be crucial in promoting safe, ethical, and unbiased AI systems. Such records should be maintained for a minimum of seven years from the time at which the AI system is no longer being publicly deployed.

We appreciate the opportunity to submit these comments and we are happy to answer any additional questions.

Respectfully submitted,

Keith Kupferschmid
CEO
Copyright Alliance
1331 F Street, NW, Suite 950
Washington, D.C. 20004

June 12, 2023

---

[9] e.g., https://contentauthenticity.org/

[10] Unless contrary to obligations under other laws, contracts, or collective bargaining agreements.

[11] Data laundering is a major issue in the AI context. Data laundering occurs when datasets comprised of copyrighted works are compiled for non-commercial research purposes, and later used for commercial purposes. In many cases, that initial research is even funded by commercial entities with the intent to use those datasets comprised of copyrighted works down the line for commercial gain. The practice of data laundering is an attempt to avoid accountability.