



**BEFORE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY**

**Request for Information on the
Current and Future State of
Emerging Technologies**

Docket No. 211116-0234

COMMENTS OF THE COPYRIGHT ALLIANCE

The Copyright Alliance appreciates the opportunity to submit the following comments in response to the *Request for Information* on the current and future state of emerging technologies published by the National Institute of Standards and Technology (NIST) in the Federal Register on November 22, 2021.

The Copyright Alliance is a non-profit, non-partisan public interest and educational organization representing the copyright interests of over 1.8 million individual creators and over 13,000 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.

In an attempt to respond to the questions posed in the Federal Register notice we solicited input from our members to develop a consensus response to the questions that represents the views of a broad section of the copyright community. What we discovered is that of the eight technologies identified by NIST in the Federal Register notice our members were most interested in and concerned about Artificial Intelligence (AI) and its potential impact on the operation of copyright law. Many of our members had differing approaches to the copyright implications of AI, due in large part to varying degrees of knowledge about and attitudes toward AI, but also due in part to the type(s) of copyright works a particular Copyright Alliance member represents and how AI

might impact those works. Despite these differing viewpoints, there is a strong consensus that AI is a profound technology that will implicate difficult questions as it relates to copyright law and these questions should be revisited in the future as AI technology continues to develop.

As an emerging and complex field of technology, which may involve both the ingestion and output of works of expression, AI is already impacting business and our daily lives and prompting some difficult questions. We must therefore begin to consider whether existing laws are sufficient to accomplish the underlying goals and purposes of U.S. copyright law. While it may be too soon, and potentially irresponsible, to attempt to *definitively* answer these questions at this time, it is a good time for NIST, in conjunction with the federal offices responsible for overseeing copyright law policy—the U.S. Patent and Trademark Office and the U.S. Copyright Office—to begin delving deeper into these issues and to begin the process for finding these answers. Ultimately, we hope that policymakers and interested parties will use this and other opportunities to become more knowledgeable about AI and its relationship to copyright law so that they are in a good position to proactively address these issues when the time is ripe, without acting prematurely in a manner that could cause more harm than good.

When discussing issues relating to AI and copyright, we believe that there are some general principles that must form the basis of any discussion. The first of these is respect for all intellectual property, especially copyright. Determining how intellectual property rules apply to a new technology is not a new or unique challenge. Whenever a new technology comes along there are some that view it as opportunity to weaken copyright. For example, when the internet was at its very nascent stages, individuals suggested that copyright should not apply online. We still hear the mantra today that “information wants to be free.” Any such view should not be tolerated and play no role in these discussions. It’s important that all policymakers and stakeholders come to the table with a common understanding about the value and importance of copyright, that the interests of those developing AI systems not be prioritized over the rights and interests of creators and copyright owners, and that the advent of AI not be used as a tool for weakening existing areas of copyright.

The second principle is the importance of education—education about AI, the value and importance of copyright, and the relationship between the two. AI is relatively new so we are just beginning to grasp the current relationship between copyright and AI and what it may be in the future. The term “AI” encompasses many technologies and use cases, so when we talk about AI it is not only important to have a common understanding of what AI is, but also to be specific about the particular applications of AI that are the source of concerns. Creating workable rules will require serious, thoughtful study of all potential ramifications and a good understanding of what AI is now and where it is going.

There needs to be efforts to educate participants in the AI industries to respect third party rights such as copyright and otherwise act in an ethical and lawful manner. Because of the disparate knowledge about AI within the copyright community, there is also a real need to educate stakeholders and policymakers about AI and its relationship to copyright. It’s difficult to develop the rules of the road when there’s not a shared understanding of how to read the map. An example of this can be found in the NOI itself. The NOI does not provide a common definition of AI that can be used by respondents to answer the questions, nor does it ask the respondents to provide their own definition.

AI is a broad and evolving field of technology, which can be categorized into generative or discriminative models, and can encompass and overlap with fields like text and data mining, and machine learning. Without a common understanding or an agreed-to definition of AI it will be difficult to determine where there may be agreement or disagreement between the stakeholders. Some respondents may have differing views and definitions regarding the scope of AI which could in turn lead to unintended misunderstandings. The first step in understanding these key AI issues is to better understand AI itself.

Of course, the United States is not the only country considering the complex issues involving AI and copyright. In particular, the European Union has been active in this area. The EU Member States joined forces with the European Commission in December 2018 to develop a Coordinated Plan on AI and in January 2019, to develop a [report](#) on European industrial policy on artificial intelligence and robotics. Japan has revised its copyright laws to address AI. Policymakers and stakeholders in the United States should take note of the analysis and decisions made by policymakers in these other countries.

The third principle is the recognition that it's important not to talk in generalities when we discuss AI and copyright. The answers to many questions that relate to the interplay between copyright and AI will need to be evaluated on their own facts and on a case-by-case basis. Given that the term "AI" encompasses many technologies and use cases, we need to be specific about the particular applications of AI that are the source of concern and to be specific about the copyright-related issues. The questions that arise and our answers to those questions should be part of an ongoing dialogue so that as AI technology and our understanding of AI changes we can better hone our thinking on the interplay between AI and copyright.

We attempt to address some of the most relevant issues below:

Copyrightability and Authorship of AI-Created Works

There are examples of AI working without human involvement closely associated, but they are few and often not well known amongst the copyright community. This is an area where education can be very helpful.

It is our present view that AI should be thought of as a tool that aids humans in creating works. The involvement of a natural person should be presupposed, and the inquiry should be whether the work qualifies as copyrightable expression, in accordance with title 17. How to identify the particular natural person or persons who legally constitute the ultimate author for copyright purposes may vary depending on the context, and in many cases will be resolved by contracts.

Ownership

Where there are questions as to ownership of an AI output, licenses and other agreements will usually be determinative. At this stage, although the factual analysis may be different, we do not think AI presents legal issues relating to ownership or transfer that are any different than those that exist today outside the area of AI. As a result, at this stage, it is not possible to make any categorical statements about ownership of the outputs created by AI.

Licensing

When discussing AI, there is an unfortunate tendency for policymakers and stakeholders to focus on the output of AI and ignore the input. In fact, the input is just as important if not more

important than the output. Some of the best training data often happens to be valuable copyrighted works. In this regard, copyright law is a huge driver of AI because it forms the basis on which high-quality training input gets created. In short, copyright owners and creators should not be expected to subsidize technology company's AI projects.

Copyright law should allow owners to exploit markets for licensing for AI uses, if they so choose. Where a copyright owner offers licenses for uses relating to the ingestion and/or training of AI system, it is essential that these licenses be respected by any copyright or AI legal regime, especially in the case of systematic ingestion of copyrighted material.

In this context, it is worth mentioning some specific concerns that certain sectors of the creative community have with generative AI systems. Generative systems and in particular Generative Adversarial Network (GANs) pose a potential risk to human creators in at least two ways. First, those systems often rely on creative content (which is often copyrighted content) as input data and are trained to make derivative works that may be sufficiently similar to the original work to fool a computer. Second, the outputs of generative AI tools often compete directly in the market with the copyrighted content on which they are trained. Licensing is key in this area since otherwise creators do not have an opportunity to be compensated for the value of their work.

Infringement

Where a particular function or output of AI implicates one or more of the exclusive rights under section 106 of the Copyright Act, copyright law should continue to protect those rights, regardless of whether the use is consumptive or non-consumptive. As with other considerations in copyright law, evaluating whether a particular AI algorithm, function or output is infringing will be a fact-specific inquiry that should be decided on a case- by-case basis. Currently, this is too complex an area for bright line rules. It is therefore essential that we differentiate between types of copying for AI functional uses. This is especially the case where AI systems generate creative content that competes directly with the copyrighted works on which they were trained.

There are some who believe that use of copyrighted works for AI ingestion is not an infringing use because it categorically qualifies as a fair use under section 107 of the Copyright Act. We strongly disagree with this view. Ultimately whether ingestion of copyrighted works is a fair use or not will depend on the facts in each particular case. While there may be instances where ingestion and training qualify as a fair use under section 107, more often than not that will not be the case. It is also important that those developing AI systems through the ingestion of copyrighted works clearly articulate the purpose of the resulting system or system output, as the purpose of the use would affect any subsequent fair use analysis.

Any fair use analysis relating to AI should carefully consider (i) how the use of copyrighted materials that are ingested by AI for training or other purposes can harm the market for these materials and the copyright owner of these materials when they are not properly licensed and (ii) how an AI system's output could affect the market for copyrighted material.

Liability

The ability to hold a natural or legal person responsible when AI infringes is an issue that will eventually need to be proactively addressed. Our initial view is that liability should attach to the beneficiary of the output of the AI—*i.e.*, the person or entity whose authorship or ownership affords them the benefits of the Copyright Act should also be responsible should that work

infringe the rights of another under the Act. When evaluating whether copying has occurred in the creation of an output for purposes of determining infringement, courts should consider not only at whether the AI ingested a particular work *at the direction of* a natural person, but also whether the AI was *enabled by* a natural person to ingest the work on its own. In many cases liability may also be addressed by a contract between the parties.

Record Keeping & Security

On thing that everyone seems to agree on is that comprehensive records of what materials, both copyrighted and not copyrighted, are used to train AI systems must be maintained and made available. From a copyright perspective, these records are important because any copyright infringement analysis would require a determination of whether “actual copying” was involved—which could be very difficult unless there is a record of the training data that was used. (Record keeping also is important to analyze for bias in the inputs.) Similar to financial records, records relating to AI and machine learning (ML) could be audited and would be discoverable during litigation. Not keeping auditable records could be treated as spoliation of evidence under discovery rules. Even in instances where content is licensed, it is still useful to keep auditable records.

Record keeping should go hand in hand with security measures and clear limits to the length of time that the user can retain the material. Where permission to use the material is granted through a license these terms will be determined by the license, but where a fair use claim is made and there is no license, courts must consider these additional elements when analyzing the fair use factors. In particular, it is essential that commercially effective security measures be employed to ensure that copyrighted material ingested by AI for training or other purposes are not compromised. Inadequate security measures may lead to unauthorized uses of the material which, among other harms, would have a detrimental effect on the copyright owner’s market for the material. Similar, retaining the material for longer than is absolutely necessary could harm the market for the material as well.

Of all the issues addressed in our comments, the standardization of record keeping and security are the areas that NIST is most qualified to pursue. If there is one takeaway from our comments on the interplay between AI and copyright it is for NIST to begin a process for standardization of record keeping and security of training material.

We thank you for the opportunity to submit these comments. Please let us know if we can provide any additional input.

Respectfully submitted,

Keith Kupferschmid
CEO
Copyright Alliance
1331 H Street, NW, Suite 701
Washington, D.C., 20005

January 31, 2022