



POSITION PAPER

Artificial Intelligence

The Copyright Alliance supports the responsible development of AI technologies and a thriving and robust artificial intelligence (AI) economy. The continuing development of AI systems represents a profound achievement of the digital age that brings with it tremendous opportunities. In fact, many in the creative industry are already using or plan to use AI for the creation of a wide range of works that benefit society. But as with many advances in technology, these new opportunities come with challenges.¹

Advancements in AI have led to a new frontier in generative technologies, and thus they are often accompanied by difficult legal questions surrounding both the ingestion of copyrighted works into AI systems and the output. As AI technology continues to evolve and questions arise about how copyright laws apply to the creation of AI-generated works, it's critical that the underlying goals and purposes of our copyright system are upheld and that the rights of creators and copyright owners are respected.

When examining the intersection of AI and copyright, the following general principles must form the basis of a common understanding amongst stakeholders, courts, policymakers, and the public.

- *When formulating new AI laws and policies, it is essential that the rights of creators and copyright owners be respected.* When making determinations about AI policies, it is vital for policymakers and stakeholders to understand that any new laws and policies relating to AI must be based on a foundation that preserves the integrity of the rights of copyright owners and their licensing markets. The interests of those using copyrighted materials to train AI must not be prioritized over the rights and interests of creators and copyright owners.
- *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.* Established copyright laws must not be weakened based on a mistaken belief that doing so is necessary to incentivize AI technologies. This is especially true when there is no evidence of market failure or problems warranting changes to the law. AI-specific statutory exceptions to copyright law that would effectively strip rightsholders of their ability to control and be compensated for the use of their copyrighted works for training purposes are not necessary and should be rejected.

¹ This paper addresses topics specific to the use of copyrighted works in training AI systems. There are several other questions that will arise as to who, if anyone, is the “author” of a work generated by an AI system, who, if anyone, is responsible for any copyright infringement committed via such system, and the copyrightability of AI-generated works in general. Those subjects will be the focus of future position papers.

- *Education is paramount in the AI space.* There must be efforts to educate participants and users in the AI industries to respect third-party rights such as copyright and otherwise act in an ethical and lawful manner.

Some of the most relevant areas of interest for the copyright community include:

Benefits of Licensing

Independent-to-large-scale creators and copyright owners produce high-quality works that are often ideal for training AI machines to generate output, and 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 uses relating to the training of AI systems, it is essential that these licenses be respected by any copyright or AI legal regime, especially in the case of ingestion of copyrighted material used for text and data mining (TDM). There is already high demand for corpuses of copyrighted works to train AI systems, and copyright owners already enter into licensing agreements for TDM use. This licensing activity is evidence of existing markets for TDM. It is important that the conditions of those licenses are respected and that they are not undermined by new exceptions that excuse unauthorized uses.

Copyrighted works are also being licensed and used for AI projects that in turn generate works that serve as market substitutes for the ingested works. In some cases, the output could qualify as derivatives of the ingested, copyrighted works. In either scenario, copyright owners and creators would be harmed from the unauthorized use of their works, and it is essential that those using the copyrighted works license such uses. In short, the marketplace should continue to properly value and incentivize creativity, and AI policy should not interfere with the ability of copyright owners to license their works for AI uses. Finally, copyright owners may sometimes choose not to license their works for use in generative systems that may produce competing output, and those choices must be respected.

Fair Use

There are some who believe that use of copyrighted works for AI ingestion will always qualify as a fair use under section 107 of the Copyright Act. That view is inaccurate. While there may be instances where ingestion and training qualify as a fair use under section 107, that likely would not be the case if a TDM license is available, the use is commercial, or the resulting AI generated work harms the actual or potential market for the ingested work. The answer will depend on the facts in each particular case.

Some AI developers have, without authorization, used training data sets or pre-trained AI created by non-commercial third parties in their commercial products—a practice known as data laundering. Neither this kind of unauthorized use nor the work of the non-commercial entity necessarily qualify as fair use. Ultimately, AI systems should only train on works or databases of content that they have the authority to use.

Transparency

Best practices from corporations, research institutions, governments, and other organizations that encourage transparency around AI training data already exist. Transparency includes such things as recording what works are ingested by AI systems and for what purpose, which helps to ensure that copyright owners' rights are respected. Infringement analyses, fair use defenses, and licensing terms disputes can all benefit from transparency best practices, and they can also be crucial in promoting safe, ethical, and unbiased AI systems.

Education & Awareness

As technologies rapidly advance, we caution against forging ahead in a way that would disregard the fundamental legal considerations at the heart of our copyright system. It is crucial that those leading AI projects are aware of the legal implications of using copyrighted works as input material, and those that arise from AI-generated output. Policymakers, in conjunction with stakeholders, must work together on educational initiatives that aim to establish common understandings and educational guidelines that ensure the rights of all are understood and respected as AI technologies evolve.