Before the
COPYRIGHT OFFICE
LIBRARY OF CONGRESS
Washington, DC

In the Matter of	)	
Artificial Intelligence and Copyright	)	Docket No. 2023-6

COMMENT OF THE COPIA INSTITUTE

I. Preliminary Statement

The Copia Institute submits the following comment to address the training questions at #9 and #14. The answer to #14, included first, speaks to the general analytical approach that should be brought to bear when considering the role of copyright law as a potentially limiting force governing the training of AI systems, and why. It should be kept in mind while reading the answer to #9, which speaks more specifically as to why copyright law does not have a role to play in limiting AI training. The comment as a whole should also be read in the context of any other question where the response may be relevant.

II. About the Copia Institute

The Copia Institute is the think tank arm of Floor64, Inc., the privately-held small business behind Techdirt.com ("Techdirt"), an online publication that has chronicled technology law and policy for over 25 years. In this time Techdirt has published more than 70,000 articles regarding subjects such as freedom of expression, platform liability, copyright, trademark, patents, privacy, innovation policy, and more. The site often receives more than a million page views per month, and its articles have attracted nearly two million reader comments, which itself is user expression that advances discovery and discussion around these topics.

As a think tank the Copia Institute also produces evidence-driven white papers examining the underpinnings of tech policy. Then, armed with its insight, it regularly files regulatory comments, amicus briefs, and other advocacy instruments on these subjects to help educate lawmakers, courts, and other regulators—as well as innovators, entrepreneurs, and the public—with the goal of influencing good policy that promotes and sustains innovation and expression. Many such filings have implicated the exact same issues as those at the fore of this inquiry.

III. Comment

A. Question #14 – Please describe any other factors you believe are relevant with respect to potential copyright liability for training AI models

Copyright is, inherently, a limitation of rights of the public. It is one that the Constitution endorses because of the resulting benefits it brings in specific, limited circumstances. However, it is meant to be limited, and we should not be willing to extend the power of copyright outside of those limitations. This is particularly true when talking about a new, dynamic, and evolving technology like "artificial intelligence."

“Artificial intelligence” has become a meaningless buzzword and one that distorts the entire discussion about its risks and merits as a technology. It is also a term that reveals little about what sort of technology it actually is – despite the nomenclature, it currently is just software, albeit sophisticated software, rather than a ready replacement for human intellect – or how its technological qualities should bear on any resulting policy discourse. Too often “AI” is a term invoked to either connote a form of magic that miraculously can solve all sorts of previously intractable problems, or a power so dangerous that its use threatens the survival of humanity.

AI, of course, is capable of both promise and peril, even in its current form today. It can also inspire both naive enthusiasm prone to deploying it in damaging ways, and also equally unfounded moral panics, preventing it from being used beneficially, as well as

genuine concerns and genuine excitement. Any policy discussion must therefore be able to tease out when fear or optimism is either misplaced or valid, and to do that it is critical to keep the discussion focused on the exact dimensions of what is being discussed when we speak of AI as a technology and not take analytical shortcuts.

It is also must recognize – as this study, to its credit, does – that AI presents a variety of different challenges to the previous technological status quo, and that each must be understood and treated separately. For instance, how we use AI, and the opportunities and potential harms such use tempts, is a fundamentally different question from how we develop or "train" AI. But too often concerns related to the former question end up shaping our policy reactions on the latter front.

This dynamic is particularly acute in the copyright context because copyright can act as such a dramatic brake to the development and use of all sorts of technologies. It is a brake that absolutely must be used judiciously and only when statutorily and constitutionally appropriate, lest it becomes an obstacle to necessary innovation, which presents its own harm. Because even if, for example, a particular use of AI might cause harm, misusing copyright law to stop it means that copyright law may also prevent a use of AI that could provide needed benefits that we now won't get.

There is also further collateral harm in bringing copyright law to bear against technology, including software-based technology such as the one generally being discussed here. The development and use of software often implicates, both directly and indirectly, expression and with it rights of free expression. Copyright inherently interferes with those rights, and consequently should not be presumed to operate here unless the statute clearly authorizes it, and the constitution allows for it. As discussed below, neither is likely when it comes to the training of AI.

B. Question #9 - Should copyright owners have to affirmatively consent (opt in) to the use of their works for training materials, or should they be provided with the means to object (opt out)?

The short answer to both these questions is no. No copyright owner should have any right to prevent their works from being used to train AI training. But the why matters.

Ultimately when we speak of training AI we are speaking about letting software "read" or otherwise consume works on behalf of the people developing it. Copyright law does not prevent reading; its exclusive rights pertain to controlling the creation (or performance) of works but not to the consumption of those created works. Nor could it, because the point of copyright law is to make it so there are works for the public to consume. It would be an absurd result – and one inconsistent with what the Progress Clause of the Constitution enables copyright law to do – if copyright law could prevent the public from getting to consume the works that copyright law has incentivized the creation of. Such barriers would also conflict with the right to read found in the First Amendment (or, stated more broadly, the right to receive information and ideas). *See, e.g.*, *Board of Education v. Pico*, 457 U.S. 853, 866-67 (1982).

In other words, the developers of an AI system would have the right to read all the works themselves. But that right is not curtailed by the use of tools – including software tools – to help them do that reading. *See Reno v. ACLU*, 521 U.S. 844, 860 (1997) (finding that the First Amendment applies to the use of computer technology to aid in the exercise of free expression). The ability to use tools to receive and perceive created works is often integral to facilitating that consumption – after all, how could the public listen to a record without a record player, or consume digital media without a computer. No law could prevent the use of tools without seriously impinging upon the inherent right to consume the works entirely. The United States is also a signatory to the Marrakesh Treaty, which addresses the unique need by those with visual and audio impairments to use tools such as screen readers to help them consume the works to which they would otherwise be entitled to perceive. Of course, it is not only those with such impairments who may have need to use such tools, and the right to format shift should allow anyone to use a screen reader to help them consume works if such tools will help them glean those ideas effectively.

AI training must be understood as simply being an extension of these same principles that allow the public to use tools, including software tools, to help them consume works. If people can direct their screen reader to read one work, they should be able to direct their screen reader to read many works. Conversely, if they *cannot* use a tool to read many works, then it undermines their ability to use a tool to help them read any. Thus it is critically important that copyright law not interfere with AI training in order not to interfere with the public's right to consume works as they currently should be able to do.

Which means, at minimum, that such AI training needs to be considered a fair use. But to say it is allowed as a fair use is to inflate the power of a copyright holder beyond what the statute or Constitution should allow because it suggests that using tools to consume works could ever potentially be an infringement. It only would not be infringement here because the AI training context would happen to excuse it. But copyright law is not supposed to give copyright owners such power over the consumption of their works, which we should be dependent on fair use to temper. It should never apply to limit the consumption of works in any context at all, and we should not let concerns about AI generally, or their uses or outputs specifically, to open the door to copyright law ever become an obstacle to that consumption. Regardless of any other consideration raised by AI, when it comes to training there is no role for copyright law to play at all.

IV. Conclusion

The First Amendment and other Constitutional and statutory constraints preclude infringement liability for AI training.

Dated: October 30, 2023

Respectfully submitted,

/s/ Catherine R. Gellis

(CA Bar # 251927)
cathy@cgcounsel.com
Counsel for the Copia Institute