

WHY U.S. COPYRIGHT LAW SUPPORTS AI INNOVATION

As debates continue around generative AI and copyright, AI Progress' new report, *AI Models: Addressing Misconceptions About Training and Copyright*, separates fact from fiction about how these AI models are trained and what the law protects. Much of the current conversation is fueled by misconceptions about how AI models use training data. And the stakes couldn't be higher: Restricting AI model development due to misinterpretations of U.S. copyright law would hinder innovation, stall progress, and undercut America's global leadership in AI.

Generative AI — especially large language models (LLMs) — are trained on large datasets to learn how to generate accurate, relevant responses. Unlike databases, these models do not need to store or systematically retrieve content.

Instead, they generate new material based on patterns, structures, and word relationships learned during training. Some critics argue this process violates U.S. copyright law. But the law protects specific creative expression, not underlying facts or patterns — which are the elements used by AI models.

U.S. copyright law — particularly the fair use doctrine — provides clear protections for this kind of transformative use. Courts have consistently recognized that innovative technologies qualify as fair use when they use copyrighted materials in new ways that don't substitute for the original works in the marketplace. Recently, two courts have found that generative AI training meets this standard.

KEY TAKEAWAYS

AI MODELS GENERATE NEW WORKS BASED ON LEARNED PATTERNS – NOT STORED DATA.



They do not need to retrieve or reproduce original training material, but instead rely on learned statistical relationships or patterns from across a broad body of content to generate responses.

COPYRIGHT LAW PROTECTS EXPRESSION, NOT IDEAS.



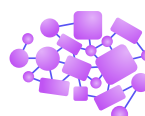
U.S. copyright law does not protect ideas, facts, or patterns revealed in material, just the specific way those things are conveyed. Similarly, the law doesn't protect against new competition — only against copying existing expression.

COURTS BACK USING COPYRIGHTED MATERIALS IN GENERATIVE AI TRAINING.



Courts have found that transformative uses — including generative AI training on whole books or datasets — are lawful when they serve a new purpose and don't compete with the original works.

EFFORTS TO RESTRICT AI TRAINING THROUGH COPYRIGHT LAW THREATEN INNOVATION.



Limiting access to training data would stall the development of AI technologies and run counter to the goals of copyright, which seeks to promote creativity and progress.