# The New York Times OpenAI Lawsuit and the Future of Machine Learning

Written by Kate Rattray

Artificial intelligence (AI) has come to play a crucial role in many lives and professions. Those using the technology place significant trust in AI platforms—yet many of us don't question where our large language models (LLMs) are getting the information they rely on to help us create AI-generated work products or complete daily tasks.

This issue—the source of LLM learning—was put front and center when the New York Times brought a federal copyright infringement lawsuit against OpenAI (the creator of ChatGPT) and Microsoft. The lawsuit alleged that OpenAI used copyrighted articles from the New York Times to create "substitutive products" without their consent. Specifically, OpenAI used, among others, New York Times content to develop their models and tools.

## Machine learning, copyright infringement, and fair use

Where does copyright infringement intersect with an AI lawsuit? It mainly has to do with machine learning, the method by which AI tools are trained to provide responses to AI prompts.

To predict the best answer to a particular question or prompt, humans need to "train" AI by feeding it information. OpenAI trains its models by feeding them large amounts of text data from online sources—including websites like the New York Times.

The problem—at least, from the New York Times's perspective—is that OpenAI and Microsoft are effectively benefitting from the New York Times's investment in journalism, using their copyrighted work to create new products without requesting permission or paying the New York Times for the service.

OpenAI and Microsoft, on the other hand, argue that their use of copyrighted content to train their models falls under "fair use", or the right to use copyrighted work without the owner's consent under certain conditions.

# Other AI lawsuits concerning copyright infringement

While the New York Times's lawsuit against OpenAI and Microsoft dominated news cycles, it's by no means the first AI lawsuit relating to copyright infringement. Other content creators—including authors like Mona Awad and Paul Tremblay, and comedian Sarah Silverman—have also initiated lawsuits against AI companies over copyright infringement.

As of the date of publication of this article, we don't have clear answers on whether machine learning qualifies as "fair use" and protects AI companies from copyright infringement lawsuits. However, the emergence of these lawsuits raises interesting questions regarding the future of AI learning.

# What the OpenAI lawsuit means for lawyers

The OpenAI lawsuit poses an interesting problem for tools designed to improve access to information: If there are restrictions on the type of information available to OpenAI and other LLMs, what are the implications for machine learning and, subsequently, the types of responses or work products these models can provide for users?

We've previously discussed the risk of bias in AI tools and the importance of ensuring that AI algorithms are developed and trained using diverse and representative data sets.

Alternatively, if LLMs can avoid copyright lawsuits by obtaining permission to use content (likely, by compensating the owners), it may follow that only those LLMs with adequate funding will have access to the comprehensive content needed to inform their models, resulting in a deleterious impact on innovation in this space.

On the other hand, if the court decides to permit LLMs to use copyrighted content for training, content creators face their own challenges—with associated consequences for the wider public. For example, content creators may limit online access for readers or viewers—to say nothing of their frustration over LLMs using their copyrighted work.

As the courts have yet to decide on the role copyright law plays in training LLMs, it's anyone's guess what the future will hold—but if plaintiffs like the New York Times are successful with their AI lawsuits, it could have a critical impact on the future of LLM training.

# The OpenAI lawsuit and our final thoughts

Regardless of the outcome, the OpenAI lawsuit is a reminder for users to carefully vet the LLMs they're using—especially when nearly one in five legal professionals currently using AI in their practice.

So, what can you do?

Before using an LLM (now, and in the wake of the court's decision), perform due diligence to understand how the tool has been trained and identify any knowledge gaps. Consider how these gaps might influence its responses. And, once you have your sights set on an LLM, remember: AI is not a replacement for your own wisdom. Carefully vet AI work products before relying on them.

*Have you heard about Manage AI? Clio's proprietary generative AI technology that is available within Clio Manage—smarter practice management starts here.*

---

## About Clio

Clio is the global leader in legal AI, powering legal work across the profession, from law firms of every size to corporate legal teams, government agencies, and courts. Its Intelligent Legal Work Platform unites the business and practice of law, connecting the people, information, and processes behind legal work. Clio's AI brings the context together with legal sources drawn from Clio Library, a growing collection of more than one billion legal documents. Over 400,000 legal professionals in more than 130 countries rely on Clio to run their legal world. Learn more at www.clio.com.