# Now & Next

## Intellectual Property Alert

May 30, 2025

# Copyright Office sets out framework for fair use of copyrighted works in AI training

**By Peter Chan[1] and Matthew A. Werber**

New fair use framework for AI training aims to balance the rights of developers and copyright holders and clarifies key factors for legal analysis.

## What's the impact?

- The Copyright Office's fair use analysis for AI training use depends on context; each case is fact-specific.

- General-purpose models using diverse data are more likely to qualify as fair use.

- Market harm, both direct loss and broader market dilution, can weigh against fair use—even without direct loss to a specific original work.

In March 2023, four months after OpenAI released ChatGPT to the public, the Copyright Office launched an AI Initiative to better understand the copyright-related legal and policy issues associated with the emergence of AI technology. In 2024 and early 2025, the Copyright Office

---

[1] Peter Chan (Legal Intern—Intellectual Property group) assisted with the preparation of this alert.

released Parts 1 and 2 of the planned multipart report, which focused on digital replicas and copyrightability. On Friday, May 9, the Copyright Office released the pre-publication version[2] of [Part 3 of the report](), which focuses on whether using unlicensed copyrighted data for AI training can qualify as fair use. AI models require massive amounts of data to train, and a significant amount of that data may be sourced from unlicensed copyrighted works, which could implicate infringement issues.[3] Therefore, the legality of training or deploying AI models may often hinge on the fair use defense.

## The Copyright Office's approach to fair use in AI development

Overall, the report strikes a balance between the rights of AI developers, on the one hand, and the rights of copyright holders, on the other hand. Instead of drawing a bright-line rule, the office concluded that fair use analysis is highly dependent on the facts and circumstances of the AI model pipeline, from initial development to the final deployment. The office applied the four "fair use" factors set forth in the Copyright Act: (1) purpose and character of use, (2) nature of the copyrighted work, (3) amount and substantiality of the portion taken, and (4) effect on market. The report focused primarily on factors 1 and 4.

## Key considerations for AI impacting fair use analysis

### FACTOR 1—PURPOSE AND CHARACTER OF USE

Factor 1 focuses on the purpose and character of the use of the copyrighted data, including whether such use is "transformative." In the AI context, this means evaluating the purpose of the training or deployment of the model. The report noted that a general foundational model, trained on diverse datasets and intended for general use, is more likely transformative fair use. For example, Meta could argue that its training of Llama would be considered transformative fair use due to its diverse training data and status as a general foundational model. On the other end of the spectrum is a domain-specific model, trained and deployed for purposes similar to the unlicensed data use. For example, a model trained solely on images from a popular animated series is less likely transformative fair use. Other factors, such as whether the deployment includes output safeguards, whether the training used pirated or illegally accessed data, or whether one or more parties in the deployment chain are commercial parties, can weigh for or against fair use.

---

[2] The released version is a pre-publication version. However, the Office does not anticipate any further substantive change to the analysis or conclusions with the final version issues in the near future.

[3] According to the report, copying can occur at various stages of training and deployment.

NIXON PEABODY

**FACTOR 4—EFFECT ON MARKET**

Factor 4 of the fair use analysis evaluates the effects of the use on the market of the copyrighted work, including the potential loss of sales or licensing revenue. Some commenters argued that courts should focus on harm to the market for the specific work used. However, the Report noted that courts may also consider the "market dilution" theory, which more generally considers whether the use harms the market for the types of data associated with the work, even if the market for the specific work is not impacted. For example, AI-generated romance novels (trained on copyrighted romance novels) may flood the market and harm the market for these types of copyrighted works, even in the absence of evidence of harm to the market for an individual romance novel.

## Practical guidance for AI developers

Overall, the report suggests a policy landscape for developers of AI models to tread confidently but carefully. Because the final analysis depends on facts and circumstances, risk-averse clients may consider licensing to reduce risk.

For more information on the content of this alert, please contact your Nixon Peabody attorney or:

**Matthew A. Werber**
312.977.4458
mwerber@nixonpeabody.com

**Erica J. Van Loon**
213.629.6031
evanloon@nixonpeabody.com

**Jason D. Hirsch**
415.984.8308
jhirsch@nixonpeabody.com

**Joshua J. Pollack**
213.629.6172
jpollack@nixonpeabody.com

**Jason T. Kunze**
312.425.3973
jkunze@nixonpeabody.com