PUBLICATIONS

# New Tech, Old Techniques: Tech Giants Are Shaping the Economic and Legal Landscape Around Social Media and Generative AI


## Related Attorneys

Stephen C. Hall

Isaac K. Keller


## Industries & Practices

Corporate Services

Intellectual Property

ARTICLE | 09.16.2026

An interesting pattern happens occasionally in law and business:

This pattern is unfolding at an extraordinary scale as generative AI developers seek massive amounts of online content to train their models. Simultaneously, social media platforms, online content creators, courts, and tech companies compete to define the rules around copying that content for training purposes.

The result is more than a copyright debate. It is a rapidly developing AI market with billions of dollars potentially at stake over the staggering amount of online content used to train these systems.

The question: how will copyright law apply to extraordinary advances in technology with current AI training methods?

## TECHNOLOGY ADVANCES: GENERATIVE AI REQUIRES MASSIVE DATASETS

Generative AI runs on content, and lots of it. The training uses massive datasets containing text, images, audio, and video, on a scale early internet users never contemplated.

Reports indicate modern AI models are trained on trillions of pages of words. If printed as books, the volume of text used to continue training a large language model for a year would dwarf even major library systems. This has produced a shift in which online information –

# New Tech, Old Techniques: Tech Giants Are Shaping the Economic and Legal Landscape Around Social Media and Generative AI

which is both incredibly abundant and freely available for viewing - is a highly used asset of the tech industry.

Indeed, an increasingly large source of training data comes from content posted by ordinary users on platforms such as Reddit, LinkedIn, X, Instagram, TikTok, and Meta's services. Unlike newspapers or stock-photo agencies, these platforms contain content generated by millions of people who likely never expected their posts, comments, photographs, and videos to be copied and used commercially. Yet now their content serves as training material for profitable AI systems.

The AI boom has transformed online content from a tool for social interaction into a valuable economic asset.

## MARKET CONDITIONS CHANGE: AI DEVELOPERS PAY FOR SOCIAL MEDIA CONTENT

When a resource becomes valuable, markets emerge.

Over several years, technology companies have entered licensing arrangements with social media platforms to access user-generated content for AI development.[1] Reddit's widely reported agreement with Google highlights the value AI developers place on continuously updated human conversation. A 2024 SEC disclosure by Reddit references "1.2 million posts created *daily* and 9.7 million comments created *daily*," as a "foundational part of how many of the leading large language models (LLMs) have been trained." (Emphasis provided.)

Reddit itself characterizes the market for content as "new and evolving rapidly" for purposes of training generative AI models and other uses. It references millions of dollars generated through "content licensing agreements executed" with a number of partners.

# New Tech, Old Techniques: Tech Giants Are Shaping the Economic and Legal Landscape Around Social Media and Generative AI

Such agreements also may demonstrate that a market for AI-training content exists in the first place. This is relevant because one of the key considerations in the copyright-fair use analysis (a safe harbor defense to copyright infringement) is whether a challenged use affects an existing or potential market for copyrighted work.

If an AI developer licenses training content from a social media platform, that alone does amount to a license from each individual creator on those sites. The distinction may be significant. Since 2024, many major social media companies updated their user agreements. LinkedIn revised its terms to create an AI-training opt-out system. Reddit's updated user agreement expressly states that its license includes "the right to use your content to train AI and machine learning models." X likewise revised its terms to include language permitting the use of user content for "training [] machine learning and artificial intelligence models."

The legal interpretation of these revisions remains open. What is clear, however, is that platforms appear to recognize the growing value of user-generated content on their sites and have taken steps to clarify that the rights transferred by individual users include copying of the content for AI training.

## COURTS APPLY LAWS TO SETTLE DISPUTES: FEDERAL COURTS HAVE BEGUN DRAWING BOUNDARIES, BUT THE LINES REMAIN UNCLEAR

Section 106 of the Copyright Act provides for exclusive copyright ownership for original works. Section 107 establishes the fair use defense, which requires courts to evaluate several factors, including the purpose and profitability of the use and its effect on the potential markets for the copyrighted work.

A recent case illustrates the importance and uncertainty of these issues. In *Kadrey v. Meta*, authors, including comedian Sarah Silverman, challenged Meta's use of copyrighted books to train its

# New Tech, Old Techniques: Tech Giants Are Shaping the Economic and Legal Landscape Around Social Media and Generative AI

Llama large language model. The court granted summary judgment in Meta's favor, largely because the plaintiffs failed to produce sufficient evidence demonstrating market harm. Importantly, the decision did not conclude that no market exists for AI-training licenses. But it did find the record insufficient to support that market-related factor.

It is worth distinguishing two separate copyright questions. This article is not focused on whether AI-generated *output* infringes upon an author's work. Instead, it focuses on whether copying copyrighted material as *input* for AI training constitutes infringement or qualifies as fair use.

Even with a number of cases on AI training and copyright law decided already, definitive answers on copying of publicly available online content for AI training do not exist. As the *Meta* Judge observed, fair use remains a "fact-specific doctrine that requires case-by-case analysis that is sensitive to new technologies and their potential consequences," and results in other cases might differ if they "have better-developed records on the market effects of the defendant's use."

Put plainly, evaluating risk of copyright liability associated with AI training remains highly dependent on specific facts, and the legal boundaries are still taking shape.

## LEGAL STRATEGY ADAPTS: STRATEGIC CONTRACTS AND POLICY ADVOCACY ARE BECOMING COMPETITIVE TOOLS

As the legal landscape develops, major technology companies are pursuing multiple strategies simultaneously.

First, AI developers continue to obtain licenses from platforms and content owners wherever possible.

# New Tech, Old Techniques: Tech Giants Are Shaping the Economic and Legal Landscape Around Social Media and Generative AI

Second, social media companies have revised user agreements to clarify and expand rights associated with user-generated content.

Finally, tech companies are actively advocating legal and policy changes favorable to AI development. Google, for example, recently published a white paper arguing that: "Using publicly available web data for training models is a transformative, non-expressive use – like an art student taking inspiration from walking through a gallery – that should remain protected under fair use in the U.S. and text-and-data-mining exceptions abroad."

Together, these efforts demonstrate that legal risk is rarely managed through litigation strategies alone. Businesses also shape outcomes through contracts, licensing arrangements, industry engagement, and policy advocacy.

## LOOKING AHEAD

The intersection of social media, generative AI, and copyright law remains a moving target. What began as a technical requirement for AI development has become a significant commercial market, a growing body of litigation, and a focal point for contractual and policy innovation.

For business leaders, the broader lesson extends well beyond AI and copyright law. Companies that recognize emerging markets, anticipate legal uncertainty, and adapt their contractual and strategic approaches accordingly will be better positioned to navigate whatever technological shift comes next.

---

[1] *See* The Copyright Alliance's amicus brief in *Kadrey v. Meta*, at pp. 4, 10 & n.6, https://copyrightalliance.org/wp-content/uploads/2025/04/Copyright-Alliance-Kadrey-Brief-4.11.25-FILED-1.pdf; *see also* Copyright Alliance, "Generative AI Licensing Isn't Just Possible, It's Essential," https://copyrightalliance.org/generative-ai-licensing/ (citing original sources and media coverage).