Authors and Publisher Giants Take on AI Over Copyright

News Team*

*Correspondence: insights@the-innovation.org

Received: May 13, 2026; Accepted: May 14, 2026; Published Online: May 14, 2026; https://doi.org/10.59717/j.tiis.2026.100003

Citation: News Team (2026). Authors and Publisher Giants Take on AI Over Copyright. *The Innvoation Insights* 1:100003.

A growing legal confrontation between the publishing industry and leading Artificial Intelligence (AI) companies intensified this week as several major publishers joined forces to sue Meta Platforms and its chief executive Mark Zuckerberg over the alleged unauthorized use of their copyrighted books and academic materials to train Llama, the company's Large Language Model (LLM).

The class-action lawsuit, filed in federal court in New York, marks one of the most significant copyright challenges by far brought against a major AI developer by the publishing sector. Plaintiffs include prominent publishing houses such as Elsevier, Hachette Book Group, Macmillan Publishers, McGraw Hill and Cengage Learning, along with fiction writer and attorney Scott Turow.

According to the complaint, Meta allegedly reproduced and distributed "millions of copyrighted works" without authorization or compensation in order to develop and improve its AI systems. The lawsuit further accuses Zuckerberg of personally approving and encouraging the practices, arguing that the company embraced a "move fast and break things" approach even when dealing with copyrighted intellectual property.

The case represents a major escalation in the increasingly contentious debate over how artificial intelligence systems are trained. Large Language Models(LLM) such as Meta's Llama rely on vast quantities of text data to generate human-like responses, but critics argue that many AI companies have built their systems using books, research papers and journalistic content obtained without permission from creators or publishers.

Figure 1. AI confronts copyright protection in publishing

Particularly significant is the participation of Elsevier, one of the world's largest scientific publishers and the publisher of influential journals including *The Lancet* and *Cell*. Industry observers note that the lawsuit is the first major AI copyright action led by large publishing houses representing both commercial and academic content providers.

Meta has rejected the allegations and said it would "fight this lawsuit aggressively", maintaining that AI training can qualify as fair use under existing copyright law. AI is driving innovation, productivity and creativity across industries, argues the company.

The lawsuit is only one of a series of its kind, part of a bigger wave of litigation against AI developers. Over the past two years, authors, media organizations and publishers have increasingly challenged the use of copyrighted material in training generative AI systems, raising fundamental questions about intellectual property, technological innovation and the future economics of creative work.

In late 2023, The New York Times filed a landmark lawsuit against OpenAI and Microsoft, alleging that the companies had used millions of its copyrighted news articles without authorization to train AI systems including ChatGPT. The newspaper argued that the AI models could reproduce portions of its reporting and threatened the economic foundation of profes-

sional journalism by diverting readers and subscription revenue. The case quickly became one of the most closely watched legal battles over whether AI training constitutes "fair use" under US copyright law.

In 2023 and 2024, a group of authors, including Sarah Silverman, filed class-action lawsuits against OpenAI and Meta, alleging that their copyrighted books were used without permission to train large language models. The plaintiffs claimed that the companies trained their systems on large-scale datasets that may have included pirated book repositories such as Library Genesis. These lawsuits have raised broader concerns within the publishing industry regarding copyright protection in generative AI training.

While many of these lawsuits remain unresolved, some AI companies have already begun seeking settlements to avoid prolonged legal battles. In 2025, Anthropic agreed to pay approximately $1.5 billion to settle a class-action lawsuit brought by several authors, including Andrea Bartz, Charles Graeber and Kirk Wallace Johnson, over the latter's allegations that the company had used copyrighted books without authorization to train its Claude AI models. The settlement was widely viewed as one of the largest copyright-related agreements in the history of the artificial intelligence industry and signaled growing pressure on AI developers to establish licensing frameworks with content creators and publishers.

Legal scholars, publishers and technology analysts remain sharply divided over whether the use of copyrighted materials to train artificial intelligence models should be considered lawful under existing copyright frameworks.

Some copyright experts argue that generative AI companies have gone far beyond traditional notions of "fair use". In several court filings submitted in support of authors suing Meta, US legal scholars contended that AI developers were not merely analyzing books and articles, but using them to build commercially valuable systems capable of competing with the original creators. They warned that allowing unrestricted use of copyrighted materials could weaken incentives for journalism, academic publishing and creative writing.

Major media organizations have also portrayed the lawsuits as a defining battle over the future economics of information. Commentators from publications including The Washington Post and The Verge have argued that the entry of large academic publishers such as Elsevier signals a widening conflict between AI developers and the global knowledge industry. Analysts note that the dispute is no longer limited to writers or journalists, but increasingly involves scientific journals, educational content and research databases that form the foundation of modern academic publishing (Figure 1).

At the same time, many technology advocates maintain that restricting AI training could severely hinder innovation. Supporters of AI companies argue that large language models do not simply reproduce copyrighted works, but transform vast amounts of text into statistical patterns that enable new forms of search, communication and productivity. Some legal analysts have compared AI training to earlier internet-era disputes involving search engines and digital indexing technologies, which courts ultimately protected under fair use doctrines.

DECLARATION OF INTERESTS

The author declares no competing interests.