

*Inside Higher Ed*

January 07, 2025

## **AI and the Struggle for Control Over Research**

For those feeling queasy about academic publishers' AI deals, Günter Waibel and Dave Hansen argue the way forward is not more restrictive licenses—it's open access.

*By Günther Waibel and Dave Hansen*

---

Taylor & Francis expected to make \$75 million from AI licensing deals for academic publications last year, and Wiley \$44 million. Oxford University Press has confirmed it is working toward similar deals.

Like many academic authors, we were troubled to learn about these lucrative contracts with AI companies for training data. While most assume that publishers struck these agreements legally, some have raised questions about whether the creators of the content should have been consulted and, perhaps, gotten a share of the money changing hands.

What gives publishers the ability to strike these deals? Most academic authors own the copyright to their work, but they often transfer those rights to the publisher at the time of publication. As a result, publishers legally and logistically control the content: They own it, and it sits behind their paywall. That creates a perfect match of supply and demand between academic publishers and AI companies, which are running out of low-hanging-fruit sources of training data.

While litigation is ongoing, AI companies have placed their bets that AI training falls within fair use, a limitation on copyright that allows for unlicensed reuse in certain circumstances, especially those that are “transformative” and reuse works for a new purpose. Nevertheless, copyright remains relevant. With dozens of fair-use AI cases winding their way through the courts, a contract with a publisher provides access to high-quality, high-volume additions to a corpus while minimizing legal risk—a particularly attractive combination for well-capitalized companies optimizing for speed rather than cost.

## **So What's an Author to Do?**

Authors troubled by the way their academic work is being exploited for profit, and chagrined about their lack of agency, are now asking, “What can I do to prevent my content from becoming grist to the mills of AI?” And they become understandably even more frustrated when the answer is *probably not much*.

Other than refraining from publishing or posting anything on the internet, there is no choice an academic author can make that will definitively hold back their content from AI training. To an author looking for a way to retreat from the brave new world of AI to a place where they can watch from the sidelines until the dust has settled, paywalled publishing may feel like the safest bet.

However, taking a step back to consider the broader implications reveals that paywalled publishing is not aligned with academic self-interest. AI is not just a shiny new product that has captured the public's imagination; it is fundamentally also a computational research methodology with profound implications for what questions we can imagine asking and how those questions can be answered.

Even if an AI future sounds dystopian, it feels intuitive that ethical and beneficial uses of AI are more likely to emerge with engagement from the academy rather than by ceding the field to commercial interest. Would you rather have an AI based on peer-reviewed literature or Reddit, YouTube and 4chan? AI will be a more powerful tool for the academy, and more likely to benefit society, if it can be trained on the most trustworthy and reliable sources of writing, including peer-reviewed scholarly literature.

## **Paywalls Are Not the Answer**

If unnerved authors retreat behind the perceived safety of a publisher's paywall, the reality is that doing so will not prevent use of their works for AI purposes. On the contrary, in signing over their copyright, authors give the same commercial publishers that already exploit academic work a monopoly over their scholarship—and as distasteful as it is, that makes the publisher's exercise of exclusive control difficult to challenge.

As a result, universities currently pay high costs for a plethora of tools powered by the

very content academic authors have created, with AI platforms being the latest in a long string of examples. By contrast, open access to scholarly literature creates a training corpus for academic computational research without publishers as a gatekeeper, enables new analyses and insights, and offers full transparency into the underlying body of source materials.

And those are, of course, just some of the benefits of open access in the specific context of AI. The fundamental value proposition of open access remains: Rather than placing scholarly content behind a paywall to be read only by those who can afford to pay, open access makes scholarly content available to anyone who will benefit from knowing.

Open access expands the public good of knowledge—the core mission universities have been founded to advance. And among the chief beneficiaries of open access are scholars and the academy itself.

### **Prioritizing the Public Good**

While it can be difficult to contemplate one's own work being used to train AI—any AI—many of our own colleagues in the academy have research agendas that rely on these same technologies, computational methods and access to a large, high-quality corpus.

Scholars depend on the very same fair-use rights as AI companies when they engage in cutting-edge computational research. Fair use is meant to be a principled rule that asks whether a given use supports the underlying goals of copyright law, which is to promote the progress of science and culture. Attempting to pick and choose who benefits from a public good inevitably erodes that public good; attempting to curtail fair use for companies that some people disfavor erodes the fair-use case for all. But fair use alone is an incomplete answer to the needs of scholars and universities.

For starters, we can only make fair use of content we can actually access. If we can agree that researchers should be able to fully deploy the latest research methodologies, if we can agree that the academy should not have to pay commercial entities to benefit from their own content, if we can agree that at the end of the day, what matters most is the unrivaled public good we create, then we need to remain steadfast in not letting others dilute or diminish that good.

The answer is not a paywall or more restrictive license. The answer is open access.

*Günter Waibel is associate vice provost in the University of California Office of the President and executive director of the [California Digital Library](#), which works to amplify the collective impact of the libraries, scholarship and resources of the 10-campus UC system. Dave Hansen, J.D., is executive director of the [Authors Alliance](#), a nonprofit organization dedicated to advancing the interests of authors who want to serve the public good by sharing their works broadly.*