Journal of Copyright in Education and Librarianship

ISSN 2473-8336 | journals.ku.edu/jcel

Volume 9, Issue 1 (2026)
12th Annual Kraemer Copyright Conference Proceedings: Copyright and the Future of Libraries

Fighting the AI Data Licensing Dystopia

Yuanxiao Xu

Yuanxiao, X. (2026). Fighting the AI data licensing dystopia. *Journal of Copyright in Education and Librarianship, 9*(1), 1-5. https://doi.org/10.17161/jcel.v9i1.24258

Fighting the AI Data Licensing Dystopia

Yuanxiao Xu
Authors Alliance

Author Note: Correspondence regarding this article should be directed to Yuanxiao Xu, xu@authorsalliance.org.

Abstract

This essay summarizes Kraemer 2025 Session 11 and examines how we have arrived at the brink of an "AI data licensing dystopia." It argues for leveraging fair use and open access to nurture human creativity and democratic oversight.

Keywords: AI, licensing, data

This article received editorial review, but the authors chose to forego peer review, which was optional for submissions consisting of conference presentations.

Fighting the AI Data Licensing Dystopia

The Illusion of Licensing as an Equitable Path Forward

As backlash grows against training AI on copyrighted materials without permission, some have promoted licensing as a fair solution.

Yet the economics of licensing deals reveal just how exclusionary they are. OpenAI now pays $16 million annually to license content from Dotdash Meredith, and that is just one of many such deals OpenAI needs for training ChatGPT. Major rightsholders like NewsCorp, Condé Nast, Axel Springer, and Reddit have all entered into opaque agreements with AI companies. Even academic publishers—such as Wiley and Taylor & Francis—are following suit, earning tens of millions by selling academic publications to AI companies. These deals are brokered behind closed doors, and shaped exclusively by dominant tech firms and powerful rightsholders.

The situation is especially dire for academic authors. These authors never profited from their works in the first place. Academic authors often pay to publish and disseminate their works. Yet their scholarship is now being sold in bulk to tech companies who commodify it without input from either the authors or the broader academic community.

Because only a handful of corporations can afford the licensing fees and legal risk involved in training generative AI, the power to shape our information ecosystem is consolidating in the hands of the few. Startups, nonprofits, and universities—institutions traditionally tasked with advancing the public interest—are effectively locked out. Even a well-funded company like Anthropic has stated that it cannot afford the licensing regime now emerging from private deals between major media conglomerates and AI companies.

Licensing for training is not a model that fosters innovation for startups, fair compensation for authors, or accountability for tech companies. Instead, it entrenches monopolistic power under the guise of legitimacy. By framing data licensing as a fair exchange, dominant players obscure the reality: this system commodifies scholarship created for the public good and sells it without author consent, compensation, or control. In effect, the licensing regime transforms a collective intellectual commons into legitimized information monopolies.

Problems with Monopoly

The AI licensing dystopia, and the monopoly it necessarily entrenches, has serious social consequences. AI systems are now deployed in critical domains such as employment, housing, finance, criminal justice, and immigration enforcement. Flawed outputs have resulted in discrimination, denial of opportunities, and even wrongful arrests. If Google, Microsoft, Meta, and Amazon become the only tech

10.17161/jcel.v9i1.24258

companies who can afford to license data for training Large Language Models (LLMs), there will be no meaningful public oversight over AI training and finetuning.

Authors will have no control over how their work is used. Researchers can't meaningfully interrogate how models were finetuned. And users have no choice but to trust outputs generated by black-box LLM systems, in everything from healthcare to criminal justice.

Other common vices of monopoly will continue to exist as well. Since the 1980s when the US stopped meaningfully enforcing antitrust law, both labor and consumers have suffered. On the labor side, productivity has steadily increased without corresponding gains in labor compensation. On the consumer side, markups have tripled. We can predict with confidence that under a monopolistic system, the benefits of AI will not be shared equitably.

Guarding Copyright's Traditional Contours

To resist the AI data licensing dystopia and the plutocracy it inevitably breeds, we must return to the bedrock principles of copyright. In *Eldred v. Ashcroft*, the Supreme Court of the United States of America recognized the importance of maintaining long-standing "traditional contours" of copyright law—namely the idea-expression dichotomy, and the right of fair use.

Copyright protects expression, not ideas or facts. And fair use safeguards the public's ability to build upon copyrighted works without permission when doing so serves new, transformative purposes. This ensures a free flow of knowledge and an abundance of creative expressions.

AI training, if it merely copies facts or structures to learn underlying patterns, does not infringe copyright. Courts have long recognized that non-expressive copying—whether for search indexing (*Perfect 10, Inc. v. Google, Inc.*), interoperability (*Sega Enterprises Ltd. v. Accolade, Inc.*), or plagiarism detection (*A.V. ex rel. Vanderhye v. iParadigms*) are fair use, and this includes when copying entire works. If an AI model does not distribute infringing outputs, it should be viewed as transformative and lawful.

But the unethical behavior of big tech companies is jeopardizing this understanding: their conduct alienates the general public and the judges alike. As a result, the traditional contours of copyright risk being reshaped not through sound legal reasoning, but through knee-jerk reaction. For example, courts and policymakers are entertaining untested legal theories like market dilution, which aim to limit AI training by redefining fair use and broadening liability in ways that serve big rightsholders' monopolistic interests rather than fostering human creativity.

Such a shift would be catastrophic. The idea-expression dichotomy and the fair use right are not just an AI-specific issue. The traditional contours of copyright

enable journalism, education, parody, criticism, and research, among many socially beneficial uses. If courts, in their zeal to punish tech companies, erode fair use and the idea-expression dichotomy, all creators—and creativity—will suffer.

Open Access as the Path Forward

Fair use jurisprudence is slow to develop. We may not get a definitive court ruling on whether AI training is fair use until 2033. But in the meantime, we can shape social norms. Just as our communities influenced the outcome of the *Authors Guild, Inc. v. Google, Inc.* and *Whyte Monkee Productions v. Netflix*, we can still shape the future if we take collective action. That includes tracking key pending cases, and speaking up on behalf of independent researchers and creators, as well as public interest.

We should also commit to open access. If fair use is uncertain, open access is well within our control. By proactively releasing high-quality scholarship under open licenses, academics can help train better, more transparent, more democratic AI models. True open-source AI models would include not only model weights, but also details about training data and finetuning protocols, thereby allowing meaningful oversight to the development process.

Our Resistance

Session 11 concluded with a quote from Cory Doctorow, who, in his book *The Internet Con*, warned of "The rise of autocrats of trade: un-elected princelings whose unaccountable whims dictate how we live, work, learn, and play" (Doctorow, 2023. p.26) This is unfortunately not science fiction.

We must do our best to resist this bleak reality. And that means rejecting licensing models that entrench monopoly. It means defending fair use and promoting open access. It means educating our peers on why the future of human authors and human creativity matter.

10.17161/jcel.v9i1.24258

References

Authors Guild v. Google, Inc., No. 13-4829 (2d Cir. 2015)
A.V. ex rel. Vanderhye v. iParadigms, L.L.C., 562 F.3d 630 (4th Cir. 2009)
Doctorow, C. (2023). *The internet con: How to seize the means of computation*. Verso.
Eldred v. Ashcroft, 537 U.S. 186 (2003)
Perfect 10, Inc. v. Google, Inc., 508 F.3d 1146 (9th Cir. 2007)
Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1992)
Whyte Monkee Productions v. Netflix, 97 F.4th 1230 (10th Cir. 2024)