

Q&A: Key issues in writers' case against OpenAI explained

September 23 2023, by Christina Pazzanese



Copyright law expert Rebecca Tushnet. Credit: Jay Mallin

There's been a lot of buzz around ChatGPT, Bard, and other generative AI tools since they burst into public view back in January. But not everyone is pleased with the rise of the chatbots. Many writers, artists,

photographers, musicians, and filmmakers say tech firms are using copyrighted work to train generative AI models.

There are now several pending lawsuits against OpenAI, the developer of ChatGPT, including [one filed](#) Tuesday in [federal district court](#) in New York by the Authors Guild on behalf of dozens of best-selling writers, including Elin Hilderbrand, Jonathan Franzen, and George R.R. Martin.

The authors say OpenAI is feeding their books into ChatGPT's large language model algorithm without consent, compensation, or attribution, in violation of U.S. copyright law. Calling it a "systematic theft on a mass scale," the guild is seeking a permanent injunction and damages for lost licensing opportunities and for making authors "unwilling accomplices" in their own future market irrelevance.

OpenAI has said the books are used only to spur innovation, not to create new works, and that that practice is lawful under the "fair use" provision of copyright law.

Rebecca Tushnet studies and teaches copyright and trademark law as the Frank Stanton Professor of the First Amendment at Harvard Law School. In a conversation with the Gazette, she talked about the authors' case against OpenAI and some of the broader legal issues around emerging tech. The interview has been edited for clarity and length.

GAZETTE: Authors claim OpenAI is 'pilfering' their books to improve ChatGPT's ability to spit out 'derivative works' in clear violation of copyright laws. Is the law clear on this issue?

TUSHNET: No. And in fact, the law in terms of using works for training or for large-scale data-mining purposes has often been held to be fair

use. The internet as we know it today, with Google and image search and Google Books, wouldn't exist if it weren't fair use to use these words and for an output that was not copying.

Now, the output, there are legitimate questions about. In theory, if you create an infringing reproduction, it's still infringing even if nobody sees it. The question is one of responsibility. Should we say, "You shouldn't make computers because they can be used to infringe"—something that copyright owners actually did think 20 years ago—or should we say, "What we have here is a tool that can be used or misused, and we should focus on curbing the misuse."

GAZETTE: How does the law protect copyrighted work?

TUSHNET: You do have rights against reproduction or the creation of derivative works, subject to limits like fair use. There are other specific limits, but fair use is the big one.

GAZETTE: Is the definition of fair use the main issue here?

TUSHNET: That's where this argument is going to go. The two questions are going to be: fair use for the training data, which I think is pretty clear under current law. And then, who's responsible for the outputs—is it the prompter or is it the existence of the tool? The law, as it exists, is reasonably settled. But the law can change. I think OpenAI has the better of the argument, but we'll see what the court thinks.

GAZETTE: Are there loopholes or carveouts in the law for particular industries, like technology?

TUSHNET: The copyright law is not like a set of rules you can fit on a page; it's several hundreds of dense pages. Are there protections for specific industries? There are a ton. There are special provisions for religious camps and farmers associations, and all sorts of stuff. But none of them are all that relevant to most of the AI issues, except insofar as there are notice and takedown regimes for cases where the output of the AI is public. Congress made special rules for internet companies to deal with the fact that the scale of the internet was so big that they couldn't treat Google like an ordinary publisher.

GAZETTE: Is current copyright law sufficient to deal with this new technological frontier?

TUSHNET: Where there is a need for guardrails, copyright is not the right way to handle it. Copyright owners, in general, have an interest in getting paid, which is not an interest in having socially beneficial output, or avoiding lies or hallucinations or anything like that. Copyright doesn't handle questions like, "How do you make sure that the AI is not defaming someone or giving you instructions on how to eat a poisonous mushroom?" The law, especially fair use, was designed to be flexible and to handle new situations. And it's done that quite well.

GAZETTE: Do you have a sense yet of how the courts may look at this issue or is it too soon to tell?

TUSHNET: A lot of times there's a tendency to say, "This is completely different from anything we've ever seen. We need a new rule." Occasionally, that's right. But a lot of times our existing principles handle it. Right now, the copyright office says if a work is generated by AI, it's not copyrightable; you need a human to do the creating. That seems, to me, to be right. That being said, you can get a human involved in tweaking an AI-generated work so that it becomes an expression of

their own creativity. And that can have a valid copyright. I can't own the forest, but if I take a nice picture of the forest, I can own a copyright in my nice picture, but only in what I did.

The problem is that the news cycle runs a lot faster than the legal cycle. It's very hard to tell in the abstract when something has changed enough that you really want to jump in with a new law and when you want to let the existing legal system handle it. People end up going back to what their prior beliefs are about various things. If you believe that big tech is fundamentally evil, you will want new rules. If you believe that big [copyright](#) owners, all they want is the money, then you will probably say, "Let the legal system handle it." And so, I think we're definitely in the too-soon-to-tell period. Right now, the training set is fair use. In terms of the output, I have no question that a clever lawyer can get an output that looks infringing. But the question is, to whom should we attribute that output?

GAZETTE: Is it premature to try to settle some of the broader legal questions before the technology has fully matured?

TUSHNET: It's a good question. The problem is things often develop very unpredictably. Thomas Edison thought that businessmen would use the phonograph to record memos and mail them to each other. That's not how it was used at all. He didn't foresee anything like the music industry that we now have. The well-known risk of regulating now is that we will write laws with the assumption that they're going to do one thing, and just completely miss the actual path of technological development, including missing things that we should have been regulating. This is why my general position is, if you care about the potential for lost jobs, we need to look to labor law and unfair competition law. Copyright is not going to help you with that. In terms of defamation, the way you deal

with that is you have a rule against defaming people. It doesn't matter whether it was generated by AI: We don't want you to do that.

Provided by Harvard Gazette

Citation: Q&A: Key issues in writers' case against OpenAI explained (2023, September 23)
retrieved 27 April 2024 from

<https://techxplore.com/news/2023-09-qa-key-issues-writers-case.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.