

AI is causing panic for authors: Now the courts are involved

October 23 2023, by Brian Contreras, Los Angeles Times



Credit: Pixabay/CC0 Public Domain

When novelist Douglas Preston first started messing around with ChatGPT, he gave the AI software a challenge: Could it write an original poem based on a character from some of his books?

"It came out with this terrific poem written in iambic pentameter," Preston recalled. The result was impressive—and concerning. "What really surprised me was how much it knew about this character; way more than it possibly could have gleaned from the internet," Preston said.

The adventure writer suspected that the chatbot had somehow absorbed his work, presumably as part of the training process by which an artificial intelligence model ingests lots of data that it then synthesizes into seemingly original content.

"That was a very disturbing feeling," Preston said, "not unlike coming home and finding that someone's been in your house and taken things."

Those worries led Preston to sign on to a proposed class action lawsuit accusing OpenAI, the developer behind ChatGPT and a major player in the growing AI industry, of copyright infringement. (OpenAI recently pursued a valuation of \$80 billion to \$90 billion.)

Preston is joined in the suit by a host of other big-name authors, including John Grisham, Jonathan Franzen, Jodi Picoult and George R.R. Martin—the notoriously slow-to-publish "Game of Thrones" author who, Preston says, joined the suit out of frustration that fans were using ChatGPT to preemptively generate the last book in his series.

OpenAI, for its part, has contended that training an AI system falls under fair use protections, especially given the extent to which AI transforms the underlying training data into something new. In an emailed statement, a spokesperson for OpenAI told The Times that the firm respects authors' rights and believes that they should "benefit from AI technology."

"We're having productive conversations with many creators around the

world, including the Authors Guild, and have been working cooperatively to understand and discuss their concerns about AI," the spokesperson said. "We're optimistic we will continue to find mutually beneficial ways to work together to help people utilize new technology in a rich content ecosystem."

Nevertheless, the publishing industry is pushing back as it reckons with a software boom that's given anyone with WiFi the power to automatically generate large reams of text. In addition to Preston's suit, various other groups of authors are pursuing their own proposed class action suits against OpenAI.

"Everybody's realizing to what extent their data, their information, their creativity, has been absorbed," said Ed Nawotka, an editor at Publishers Weekly. There is, in the industry, a degree of "abject panic."

In one recent pair of lawsuits, Sarah Silverman accused OpenAI as well as Meta—Facebook's parent company and a major AI developer itself—of [copyright infringement](#). The two companies have since pushed to get most of Silverman's cases dismissed.

A different suit recently found Paul Tremblay ("The Cabin at the End of the World") and Mona Awad ("Bunny") suing OpenAI for [copyright violations](#)—the company is trying to get that one mostly dismissed too—while Michael Chabon ("The Yiddish Policemen's Union") is a plaintiff in two additional legal actions that are targeting OpenAI and Meta, respectively.

And this past July, the Authors Guild—a professional trade group, not a labor union—sent several [tech companies](#) an [open letter](#) calling for consent, credit and fair compensation when writers' works are used to train AI models. Among the signatories were Margaret Atwood, Dan Brown, James Patterson, Suzanne Collins, Roxane Gay and Celeste Ng.

That's all on top of the nearly 5-month-long strike that Hollywood screenwriters recently undertook that led to, among other things, new regulations on the use of AI for script generation. (A separate strike, still ongoing, has found screen actors rallying around AI concerns of their own.)

The lawsuit in which Preston is involved, which features 17 other named plaintiffs including the Authors Guild, claims that OpenAI copied the authors' works "without permission or consideration" in order to train AI programs that now compete with those authors for readers' time and money.

The suit also takes issue with ChatGPT's generation of derivative works, or "material that is based on, mimics, summarizes, or paraphrases [the] Plaintiffs' works, and harms the market for them."

The plaintiffs are seeking damages for their lost licensing opportunities and "market usurpation," as well as an injunction against future such practices, on behalf of American fiction authors whose copyrighted works were used to train OpenAI software.

"They didn't ask our permission, and they aren't compensating us," Preston said of OpenAI. "What they've done is created a very valuable commercial product which can reproduce our voices. ... It's basically theft of our creative work on a grand scale."

Since the plaintiffs' books aren't freely available on the open web, he added, OpenAI "almost certainly" accessed them via alleged piracy sites such as the file-sharing platform LibGen. (The suit reiterates this suspicion, attributing it to "independent AI researchers.")

OpenAI declines to answer a question about whether the plaintiffs' books were part of ChatGPT's training data or accessed via file-sharing

sites such as LibGen. In a statement to the U.S. Patent and Trademark Office cited in the Authors Guild suit, OpenAI stated that modern AI systems are sometimes trained on publicly available data sets that include copyrighted works.

The Atlantic has reported that Meta, meanwhile, trained its ChatGPT competitor LLaMA on a corpus of pirated ebooks known as "Books3." A searchable version of that data set indicates that LLaMA fed on books written by almost all of the individuals named as plaintiffs in the various aforementioned lawsuits.

The works of L.A. Times staffers were included too. Meta did not respond to a request for comment from The Times about how LLaMA was trained.

The specific sources of the training data aside, many authors are worried about where this technology is leading their industry.

Michael Connelly, the author of the Harry Bosch series of crime novels and another plaintiff in the Authors Guild lawsuit, framed those concerns as a matter of control: "control of your own work, your own property."

Connelly never got to decide whether his books would be used to train an AI, he said, but if he'd been asked—even if there were money on the table—he likely would've opted out. The idea of ChatGPT writing an unofficial Bosch sequel strikes him as a violation; even when Amazon adapted the series into a TV show, he says, he had some control over the scripts and casting.

"These characters belong to us," Connelly said. "They come out of our heads. I even put stuff in my will about [how] no other author can carry the Harry Bosch torch after I'm gone. He's mine, and I don't want anyone

else telling his story. I certainly don't want a machine telling it."

But whether the law will allow the machines to do so is a different question.

The various lawsuits against OpenAI allege copyright violations. But copyright law—and especially fair use, the area of law governing when copyrighted work can be incorporated into other endeavors, such as for the sake of education or criticism—still doesn't offer a cut-and-dry answer as to how these lawsuits will shake out.

"We've got kind of a push and pull right now in the case law," said intellectual property attorney Lance Koonce, a partner at the law firm Klaris, pointing to two recent Supreme Court cases that offer competing models of fair use.

In one, *Authors Guild vs. Google*, the court held that Google was allowed to digitize millions of copyrighted books in order to make them searchable. In the other, *Andy Warhol Foundation for the Visual Arts Inc. vs. Goldsmith*, the court found that the titular pop artist's incorporation of a photographer's work into his own art didn't fall under fair use because Warhol's art was commercial and had the same basic purpose as the original photo.

"These AI cases—and especially the Authors Guild case (against OpenAI)—fall into that tension," Koonce said.

In its patent office statement, OpenAI argued that training artificial intelligence software on copyrighted works "should not, by itself, harm the market for or value of copyrighted works" because the works are being consumed by software rather than real people.

Outside of legal avenues, stakeholders are already pitching solutions to

this tension.

Suman Kanuganti, the chief executive of AI messaging platform Personal.ai, said the tech industry will likely adopt some sort of attribution standard that allows people who contribute to an AI's training data to be identified and compensated.

"Once you build the models with known, authenticated data units, then technologically, it's not a challenge," Kanuganti said. "And once you solve that problem ... the economic association then becomes easier."

Preston, the adventure novelist, agreed that there may yet be a path forward.

Licensing books to software developers through a centralized clearing house could provide authors with a new income stream while also securing high-quality training data for AI companies, he said, adding that the Authors Guild tried to set up such an arrangement with OpenAI at one point but that the two sides were unable to reach an agreement. (OpenAI declined to discuss such conversations.)

"We were trying to get them to sit down with us in good faith; we're not opposed at all to AI," Preston said. "It's not a zero-sum game."

2023 Los Angeles Times. Distributed by Tribune Content Agency, LLC

Citation: AI is causing panic for authors: Now the courts are involved (2023, October 23)
retrieved 8 May 2024 from

<https://techxplore.com/news/2023-10-ai-panic-authors-courts-involved.html>

<p>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.</p>
--