

IP Alert | Authors' Copyright Battle Against OpenAI Survives Motion to Dismiss

By Azuka C. Dike, Rebecca Ding, and Victoria Webb

This week, Judge Sidney Stein in the Southern District of New York issued two Orders rejecting attempts to dismiss or strike claims in the consolidated multi-district class action that numerous plaintiff authors brought against OpenAI and Microsoft. The action, brought by some of the most prominent names in publishing, is part of a wave of copyright infringement lawsuits against artificial intelligence (“AI”) companies involving unauthorized use of copyrighted material to train large language models (“LLMs”) and the potential creation of infringing works in the outputs of those trained-LLM products, including in this case, outputs from OpenAI’s ChatGPT.

First, on Monday, Judge Stein denied OpenAI’s motion to dismiss a claim of direct copyright infringement brought by a consolidated class of plaintiff authors who accuse the artificial intelligence (“AI”) company’s generative AI chatbot, ChatGPT, for outputting content that infringes their copyrighted works. The decision now allows the plaintiffs to proceed with their direct infringement claim against OpenAI. Second, on Tuesday, Judge Stein issued an opinion that largely denied OpenAI’s and Microsoft’s motions to strike other claims in the Consolidated Class Action Complaint.

These decisions allow several claims to proceed, with many important issues, such as fair use, to be resolved.

The Wave of Copyright Cases Against OpenAI and Microsoft

Beginning in 2023, multiple authors began suing OpenAI and its investor Microsoft for infringing the authors’ copyrights. The wave of copyright infringement lawsuits continued to expand, with numerous cases in the Southern District of New York and more in the Northern District of

California. Eventually, after much procedure, the consolidated multi-district class action combined two previously consolidated cases in the Southern District of New York with multiple cases transferred from the Northern District of California. That consolidated Multidistrict Litigation (MDL) No. 3143 is now before Judge Sidney H. Stein in the Southern District of New York, with Magistrate Judge Ona T. Wang overseeing discovery.

Plaintiffs include some of the most prominent names in publishing and media: The New York Times Company; The Authors Guild and leading authors such as George R.R. Martin and John Grisham; and Ziff Davis, Inc., the parent company of IGN, Mashable, and CNET. Additional publishers and authors continue to join, underscoring the broad reach of the litigation.

At the core of the consolidated case are copyright infringement claims. Plaintiffs allege that OpenAI directly copied vast quantities of copyright protected works without authorization to train its LLM. Plaintiffs further allege that the trained-LLM products, such as OpenAI's ChatGPT bot, generate outputs that allegedly reproduce verbatim or derivative infringing versions of the authors' original works. Plaintiffs also assert other claims, such as Digital Millennium Copyright Act ("DMCA") violations for stripping copyright management information, trademark dilution, and unjust enrichment for allegedly avoiding licensing obligations while profiting from copyrighted works.

OpenAI Asked the Court to Dismiss Plaintiffs' Claim of Direct Infringement Related to the Allegedly-Infringing Outputs

OpenAI filed its motion to dismiss on July 14, 2025. In the motion, OpenAI asked the Court to dismiss one of plaintiffs' numerous claims—their direct copyright infringement claim based on outputs of ChatGPT—arguing that the claim fails to the extent it depends on ChatGPT outputs alleged to reproduce or otherwise infringe the plaintiffs' works. Among other things, OpenAI argued in the motion that the complaint did “not attach or even quote a single sentence from any allegedly infringing output; nor does it provide any meaningful description of what the outputs contain, much less how their contents are substantially similar to any protected expression in plaintiffs' works.”

The plaintiffs responded in an August 2025 opposition. Among other things, plaintiffs contended that the Complaint adequately alleged substantial similarity between the ChatGPT-generated outputs (e.g., summaries of the plaintiffs' original works or outlines of potential sequels) and plaintiffs' original works. Specifically, plaintiffs argued that the Complaint described with specificity the protected elements (e.g., characters, plot, narrative arcs) that ChatGPT had generated and provided allegations of plot- and character-level similarity sufficient to state a viable claim.

The Court held a hearing on October 8, 2025. At oral argument, OpenAI argued that plaintiffs' copyright allegations were directed to the scraping and copying of copyrighted works on the Internet, not that ChatGPT's summaries and prompt-generated outputs constitute additional

copyright infringement. OpenAI also raised concerns that plaintiffs' substantial similarity claims would interfere with ongoing discovery efforts. During the hearing, Judge Stein questioned OpenAI's attorneys regarding ChatGPT-generated summaries of plaintiffs' works, including George R.R. Martin's "A Song of Ice and Fire" series of fantasy novels, which were later adapted into the hit HBO television series "Game of Thrones."

The Court Denied OpenAI's Motion to Dismiss, Finding Plaintiffs "adequately state[] a prima facie claim of copyright infringement based on ChatGPT's outputs"

According to the Court's Order, the plaintiffs' Complaint provided adequate evidence of infringements by the AI output at the pleading stage. Judge Stein wrote, for example, that "[a] more discerning observer could reasonably conclude that the allegedly infringing outputs are substantially similar to plaintiffs' copyrighted works."

The Court analyzed two types of works produced by ChatGPT, the ChatGPT-generated summaries of plaintiffs' works, and the outlines for potential sequels to plaintiffs' works. The Court noted that while the exemplary summaries submitted by plaintiffs "do not recount '[e]very intricate plot twist and element of character development' in the original works," they are certainly attempts at summarizing central copyrightable elements of the original works (e.g., setting, plot, characters) and include many specific details drawn from the original works.

As an illustrative example, the Court analyzed an output from ChatGPT summarizing "A Game of Thrones" from George R. R. Martin's "A Song of Ice and Fire" book series. The generated output described the story's setting ("The Seven Kingdoms of Westeros are ruled from the Iron Throne in the capital city, King's Landing."), provided a detailed prologue ("Members of the Night's Watch, a sworn brotherhood tasked with defending the realm from threats beyond the Wall ... are attacked by mysterious and deadly creatures known as the White Walkers"), and listed several main plot points ("Stark Family in Winterfell;" "Daenerys Targaryen in Essos;" "The North and Beyond the Wall"), while also identifying multiple key characters and locations (Eddard Stark, King Robert Baratheon, Khal Drogo, Jamie Lannister, Tyrion Lannister, Winterfell, King's Landing). The Court found that the detailed summary generated by ChatGPT is substantially similar to Martin's original work because the summary "conveys the overall tone and feel of the original work by parroting the plot characters, and themes of the original."

The Court distinguished ChatGPT's summaries of Martin's books in this class action from its summaries of news articles in a separate copyright action, which were determined to be non-infringing. The Court noted that in the other action, the generated outputs merely summarized non-copyrightable elements of the original news articles (i.e., the facts contained in the article) and differed in style, tone, and length from the original article. By contrast, the Court found that the

generated summaries of Martin's books incorporated copyrightable elements of the original works, including plot, setting, and characters.

Next, the Court addressed the outlines for potential sequels to plaintiffs' works that were generated by ChatGPT. The Court similarly found that a reasonable jury applying the more discerning observer test could conclude that the derivative sequels generated by ChatGPT are substantially similar to plaintiffs' original works. An example of such an output considered by the Court is an outline of an alternative sequel to one of Martin's books ("A Clash of Kings"). The sequel, aptly titled "A Dance with Shadows," describes a distant relative of the Targaryens, Lady Elra, who raises an army in Essos and seeks to claim the Iron Throne for herself. In this reimagined version of events, Robb Stark creates a surprise alliance with Renly Baratheon to change the balance of power in Westeros, and Bran Stark heads South (not North) to warn his family of the threat from White Walkers. The Court concluded, again, that a reasonable jury applying the more discerning observer test could determine that this output is substantially similar to Martin's original work based on the output's incorporation of setting, plot, and characters from Martin's books.

Concerning the issue of whether plaintiffs provided sufficient evidence in the Complaint to prove its claims of copyright infringement, the Court concluded that the example outputs submitted by plaintiffs, in connection with their opposition to OpenAI's motion to dismiss, were incorporated into the Complaint by reference. As such, the Court determined that the example outputs could be considered for purposes of analyzing similarity with plaintiffs' works.

The Court also Largely Denies Defendants' Additional Motions to Strike: the Court Narrows Scope of GPT Models but Preserves Core Infringement Allegations

This week the Court also resolved several other motions to strike portions of the Consolidated Class Action Complaint. Judge Stein provided a small win for Defendants and granted OpenAI's motion to strike allegations related to newer and as-yet unreleased GPT models—including GPT-4V, GPT-4.5, GPT-5, and their "derivatives" or "successors"—because they exceeded the Court's prior directive limiting the consolidated case to GPT-3 through GPT-4o Mini.

Microsoft also separately sought to strike allegations regarding certain GPT models. Similar to OpenAI's motion, the Court granted part of Microsoft's motion and struck allegations regarding the same models: GPT-4V, GPT-4.5, GPT-5 and their derivatives or successors. But the Court found, for example, that "Microsoft slices the baloney too thin" in trying to strike other models such as GPT-4o and GPT-4o Mini, and denied Microsoft's other requests to strike.

The Court also rejected OpenAI's motion to strike the so-called "download claim" for copyright infringement. As the Court noted, OpenAI contended that the Complaint added new claims because it did not link together certain factual allegations. The Court noted that prior complaints

adequately put OpenAI on notice of plaintiffs' claims against it for copyright infringement based on alleged facts related to the downloading and reproduction of books. Accordingly, the Court concluded that plaintiffs' download claim was not a new or expanded cause of action. So once again, OpenAI failed at eliminating one of plaintiffs' claims.

So the Case Now Proceeds . . . With Much Left to Be Decided

The case now continues forward concerning direct infringement and other claims and defenses. While this is an initial win for the plaintiff authors, it is important to remember that the decision at the motion to dismiss stage evaluates the complaint with respect to **plausibility pleading standards** set forth in *Twombly* and *Iqbal*. Plaintiffs are not required to prove infringement at this stage, but merely to allege enough facts to make the alleged infringement plausible, i.e., that the ChatGPT-generated outputs are substantially similar to their copyrighted works. So whether the facts can ultimately carry the plaintiff authors all the way to a win is yet to be seen.

Further, Judge Stein also made clear that the question of whether the allegedly infringing ChatGPT outputs are protected as **fair use** remains unsettled in this case. And so far, **only a few District Courts have weighed in on the novel questions of fair use in AI training and generation context**. So fair use in the AI context very much remains in flux, with outcomes turning on subtle distinctions between transformative learning and unauthorized replication. Because fair use in the AI training context remains an open question here, any future consideration of fair use in the OpenAI case will help provide further direction on this issue.

Finally, given that the question of fair use remains open and unsettled, as in other AI training cases, the use of large quantities of training materials creates immense exposure for companies training LLMs if even a fraction of those works are found to be infringing. Here, the plaintiffs in the OpenAI case allege in their Complaint that Defendants "made an intentional decision to use pirated libraries" and trained their models on "at least hundreds of thousands of books." Because statutory damages could reach up to **\$150,000 per work**, OpenAI remains exposed to billions of dollars in potential liability given the vast number of works alleged to have been infringed. Given these risks and the open legal questions, AI companies and others training machine learning models should continue to carefully ensure their training data is legally acquired.

Posted: October 30, 2025