Artists, Writers Sue AI Companies for Copyright Infringement

Several lawsuits have been filed against artificial intelligence (AI) companies by artists and writers who claim that the companies have violated their copyright by using their work to train their AI models without permission or license.

On Jan. 13, 2023, three visual artists, Sarah Andersen, Karla Ortiz and Kelly McKernan, filed a class action lawsuit in federal court in California against Stability AI, Midjourney, and DeviantArt over allegations that the companies used the artists' work to train their AI models without permission, constituting copyright infringement. Each of these AI models can generate artistic work based on text prompts. The full text of the complaint is available online at: https://admin.bakerlaw.com/wp-content/uploads/2023/09/ECF-1-Complaint-2.pdf.

After several rounds of pleadings, Judge William Orrick of the U.S. District Court for the Northern District of California granted the AI companies' motion to dismiss on Oct. 30, 2023. Each of the claims brought by the plaintiffs, including direct copyright infringement, vicarious copyright infringement, violation of the Digital Millennium Copyright Act (DMCA), and violation of California's statutory and common law right to publicity, were dismissed. According to the U.S. Copyright Office, vicarious copyright infringement occurs when the party had "(1) the right and ability to supervise or control the infringing activity; and (2) a direct financial benefit from that activity." A full overview of this concept is available online at: https://www.copyright.gov/docs/regstat072204.html. The only claim to survive was Andersen's direct copyright infringement claim against Stability AI. Judge Orrick granted the plaintiffs leave to amend their complaint within 30 days to keep their claims alive.

Judge Orrick categorized the original complaint as "defective in numerous respects," ruling that the plaintiffs generally needed to be more specific about how each of the defendants "separately violated their copyrights, removed or altered their copyright management information, or violated their rights of publicity." The full text of Judge Orrick's order granting

the motion to dismiss is available online at: https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.117.0_3.pdf.

On November 29, the plaintiffs filed an amended complaint, which also added several more artists to their claim.

"Though Defendants like to describe their AI image products in lofty terms, the reality is grubbier and nastier: AI image products are primarily valued as copyright-laundering devices, promising customers the benefits of art without the costs of artists," the artists wrote in their amended complaint. The full text of the amended complaint is available online at: https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.129.0_1.pdf.

Getty Images brought a separate case against Stability AI on Feb. 3, 2023, that alleged copyright and trademark infringement. Getty Images accused Stability AI of copying "more than 12 million photographs" and their "associated captions and metadata" in an effort to build "a competing business." In its complaint, filed in the U.S. District Court for the District of Delaware, Getty Images provided examples of Stability AI's model, Stable Diffusion, which produced modified versions of watermarked Getty images that include distortions of the Getty Images' logo. The full text of the complaint is available online at: https://admin.bakerlaw.com/wp-content/uploads/2023/09/ECF-1-Complaint-1.pdf.No major developments have occurred in the case.

In addition to claims by visual artists, several writers have brought lawsuits against AI companies for copyright infringement. On June 28, 2023, authors Paul Tremblay and Mona Awad filed a class action against OpenAI for copyright infringement. On July 7, 2023, comedian Sarah Silverman and novelists Christopher Golden and Richard Kadrey filed class action lawsuits against OpenAI and Meta over similar allegations of copyright infringement. The lawsuits, which were filed in the U.S. District Court for the Northern District of California, allege that the named company copied the texts from the authors' books without permission by scraping the text from illegal "shadow libraries" and used those texts to train its AI model. The complaint against Meta is available online at: https://llmlitigation.com/pdf/03417/kadrey-meta-complaint.pdf. Golden, Kadrey, and Silverman's complaint against OpenAI is available online at: https://llmlitigation.com/pdf/03416/silverman-openai-complaint.pdf. Awad

and Tremblay's complaint against OpenAI is available online at: https://admin.bakerlaw.com/wp-content/uploads/2023/09/ECF-1-Complaint-3.pdf.

On July 28, a California district judge ordered the Tremblay and Silverman cases to be related. According to the law firm Tyson & Mendes, relatedness is a different concept than consolidation. Unlike consolidation, related cases remain separate for trial purposes, but parties should be served with all pleadings and discovery from each of the related cases. Thus, the effect of cases being "related" is primarily procedural in nature. A full overview of the difference between relatedness and consolidation is available online at: https://www.tysonmendes.com/lost-translation-demystifying-consolidation-multi-party-litigation.

On August 11, Awad voluntarily dismissed her claim against OpenAI with prejudice, offering no reason for doing so, but this did not affect the claims of the other authors in either the Tremblay or Silverman cases. On August 28, OpenAI filed motions to dismiss both the Tremblay and Silverman cases. The authors opposed the motions, but the district court has yet to rule on the motions to dismiss. A separate class action was later filed against OpenAI on September 8 in the Northern District of California by writers Michael Chabon, David Henry Hwang, Matthew Klam, Rachel Louise Snyder, and Ayelet Waldman on similar claims of copyright infringement. On October 10, the court granted a motion to relate the Tremblay and Chabon cases against OpenAI. A landing page outlining the recent updates in the Tremblay and Silverman cases against OpenAI is available online at: https://www.bakerlaw.com/tremblay-silverman-v-openai/. The full text of the Chabon complaint against OpenAI is available at: https://admin.bakerlaw.com/wp-content/uploads/2023/09/ECF-1-Chabon-OpenAI-Complaint.pdf.

In the authors' case against Meta, Judge Vince Chhabria granted Meta's motion to dismiss all but one of the claims against Meta on Nov. 20, 2023. Judge Chhabria's order left the authors' main claim of copyright infringement intact but dismissed remaining claims that alleged vicarious copyright infringement, unfair competition, unjust enrichment, and negligence. However, the judge gave the plaintiffs leave to amend their complaint by December 11 to keep each of these claims (apart from the negligence claim which was dismissed with prejudice) alive. Judge Chhabria's order granting Meta's motion to dismiss is available online at:

https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.56.0_1.pdf.

On Sept. 19, 2023, another lawsuit was filed against OpenAI, this time by the Authors Guild and other writers including John Grisham and George R. R. Martin. The complaint, which was filed in U.S. District Court for the Southern District of New York, accuses OpenAI of knowingly and wrongfully copying the writers' work in order to train its AI model, ChatGPT, in an act of "systematic theft on a mass scale." The full text of the complaint is available online at: https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.1.0_1.pdf.

The plaintiffs later amended their complaint on Dec. 4, 2023, to add Microsoft as a defendant, accusing the company of materially or directly assisting in OpenAI's alleged copyright infringement. "OpenAI's 'training' [of] its [large language models] could not have happened without Microsoft's financial and technical support," the amended complaint said. The full text of the amended complaint is available online at: https://storage.courtlistener.com/recap/gov.uscourts.nysd.606655/gov.uscourts.nysd.606655.39.0.pdf.

In response to this and the many other copyright lawsuits against it, an OpenAI spokesperson told *Reuters* on Sept. 21, 2023, that the company respects authors' rights and is "having productive conversations with many creators around the world, including the Authors Guild." The *Reuters* story is available online at: https://www.reuters.com/legal/john-grisham-other-top-us-authors-sue-openai-over-copyrights-2023-09-20/.

The Authors Guild lawsuit follows a previous attempt by the Guild to voice their grievances to AI companies regarding the issue of copyright infringement. In July 2023, the Guild and more than 15,000 writers sent an open letter to the CEOs of prominent AI companies to call attention "to the inherent injustice in exploiting our works as part of your AI systems without our consent, credit, or compensation." The full text of the authors' open letter is available online at: https://authorsguild.org/app/uploads/2023/10/Authors-Guild-Open-Letter-to-Generative-AI-Leaders.pdf.

In another example of copyright claims against AI companies, a group of six authors, led by Mike Huckabee, a former governor of Arkansas who sought the Republican nomination for president in 2008 and 2016, filed a class action

suit against Meta, Bloomberg, Microsoft, and The EleutherAI Institute, over allegations that Meta's, Bloomberg's, and Microsoft's use of EleutherAI's Books3 dataset in training their AI models infringed upon 183,000 copyrighted e-books. The complaint, which was filed on Oct. 17, 2023, in the U.S. District Court for the Southern District of New York, alleged that the defendant's wrongfully used the dataset with the "full knowledge and understanding" that the training data being used was "assembled from copyrighted works, including copyrighted works of the Plaintiffs." The case has had no significant developments since its filing. The full text of the complaint is available online at:
https://admin.bakerlaw.com/wp-content/uploads/2023/10/ECF-1-Complaint.pdf.

As the legal battles against AI companies continue, some scholars and commentators claim that copyright infringement is not the claim on which artists, writers, and other content creators should be focused. Matthew Sag, a law professor at Emory University, said that the general theory that large language models (LLMs) "copy" artists' work is misplaced.

In testimony to the U.S. Subcommittee on Intellectual Property in July 2023, Sag said, "Rather than thinking of an LLM as copying the training data like a scribe in a monastery, it makes more sense to think of it as learning from the training data like a student. If an LLM like [OpenAI's] GPT-3 is working as intended, it does not copy the training data at all. The only copying that takes place is when the training corpus is assembled and pre-processed." The full text of Sag's testimony is available online at: https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf.

"There are a lot of questions that AI creates for almost every aspect of society," Mike Masnick, editor of the technology blog *Techdirt*, told *Wired*. "But this narrow focus on copyright as the tool to deal with it, I think, is really misplaced." The full *Wired* article is available online at: https://www.wired.com/story/artificial-intelligence-copyright-law/.

Even when looking at the claim that AI companies use illegally obtained "shadow libraries" or datasets to train their LLMs, Masnick said that lawfully obtaining access to works may be irrelevant to a defense that the AI companies were merely "inspired" by the works.

"If a musician were inspired to create music in a certain genre after hearing pirated songs in that genre, would that make the songs they created

[copyright infringement]?" Masnick posited in an article on his blog *Techdirt*, which is available online at: https://www.techdirt.com/2023/07/11/a-bunch-of-authors-sue-openai-claiming-copyright-infringement-because-they-dont-understand-copyright/.

However, Karla Ortiz, who is one of the plaintiffs in the lawsuit against Stability AI, took issue with this sentiment, arguing that inspiration for humans is much different than inspiration for AI.

"Artists look at other artists to learn how to solve visual problems and be inspired, but that's as far as looking at other artists will get you. Artists bring their own technical knowledge, problem solving, experience, thoughts and personal lives into each artwork. This is also why humans who are trained in the same way, who paint the same objects will still yield different results. AI's can never accomplish that," Ortiz wrote in testimony before the U.S. Subcommittee on Intellectual Property in July 2023. The full text of Ortiz's testimony is available online at: https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_ortiz.pdf.

One of the major questions raised by these copyright cases is whether the AI companies' use of others' work is "fair use." For example, in response to the artists' suit against Stability AI, a spokesperson for the company told *Reuters* that "anyone that believes that this isn't fair use does not understand the technology and misunderstands the law." The full text of the *Reuters* article is available online at: https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/.

The U.S. Copyright Office defines fair use as "a legal doctrine that promotes freedom of expression by permitting the unlicensed use of copyright-protected works in certain circumstances." Section 107 of the Copyright Act offers four factors to consider when determining whether something constitutes fair use, including the (1) purpose and character of the use, (2) nature of the copyrighted work, (3) amount of the portion used in relation to the copyrighted work as a whole, and (4) effect of the use upon the potential market for or value of the copyrighted work. 17 U.S.C. § 107. A full overview of the fair use doctrine provided by the U.S. Copyright Office is available online at: https://www.copyright.gov/fair-use/.

Artists, writers, and other scholars disagree with the AI companies' defense that use of others' work constitutes fair use. Daniel Gervais, a

professor at Vanderbilt Law School, told *The Verge*, "If you give an AI 10 Stephen King novels and say, 'Produce a Stephen King novel,' then you're directly competing with Stephen King. Would that be fair use? Probably not."

Gervais clarified that "'it is much more likely than not' that *training* systems on copyrighted data will be covered by fair use. But the same cannot necessarily be said for generating content."

Yet, given the open issue of copyright infringement by AI models, Andy Baio, a technologist who has been closely following the generative AI scene, told *The Verge* that it is impossible at this stage to know how the courts will decide.

"I see people on both sides of this extremely confident in their positions, but the reality is nobody knows. And anyone who says they know confidently how this will play out in court is wrong," Baio said. The full text of *The Verge* article is available online at: https://www.theverge.com/23444685/generative-ai-copyright-infringement-legal-fair-use-training-data.

— Charlotte Higgins
Silha Center Research Assistant