February 6, 2024

If you have any questions regarding the matters discussed in this memorandum, please contact the following attorneys or call your regular Skadden contact.

Stuart D. Levi
Partner / New York
212.735.2750
stuart.levi@skadden.com

Mana Ghaemmaghami
Associate / New York
212.735.2594
mana.ghaemmaghami@skadden.com

Shannon N. Morgan
Associate / New York
212.735.3711
shannon.morgan@skadden.com

One Manhattan West
New York, NY 10001
212.735.3000

Motion To Dismiss Ruling Provides Insight Into How Courts View AI Training Data Cases

A recent decision by a California district court in _J. Doe 1 v. GitHub, Inc._, a case brought by computer programmers alleging that their works had been used to train AI models that generate computer code in violation of their rights, highlights how courts are grappling with some of the unique legal issues presented by these fact patterns.

Background

Original Complaint

In their original complaint filed in November 2022, two developers, J. Doe 1 and J. Doe 2, alleged that Copilot and Codex, two AI products designed to generate computer code in response to text prompts, were trained on the plaintiffs' copyrighted computer code in violation of their rights. The original complaint named as defendants GitHub (an open-source platform owned by Microsoft that the plaintiffs use to publish the code at issue and that distributes Copilot); Microsoft (as the owner of GitHub); and various OpenAI entities that programmed, trained and maintain Codex. Because the plaintiffs' code was freely available under broad open-source licenses, they could not allege some of the direct copyright infringement claims raised in other training data cases. Rather, the plaintiffs' complaint centered on other claims, including:

- Violation of the Digital Millennium Copyright Act (DMCA), for removal or alteration of copyright management information (CMI) from the plaintiffs' code.

- Breach of contract, because the applicable open-source licenses required a statement of attribution, among other things, that were not included in the models or their outputs.

- Violation of the California Consumer Privacy Act.

- Tortious interference in the plaintiffs' contractual relationships.

The defendants moved to dismiss the complaint.[1]

First Motion To Dismiss Decision

In its May 2023 ruling on the defendants' motions to dismiss, a California district court addressed various issues including Article III standing requirements, the use of pseudonyms by the plaintiffs — who did not want to be named for fear about their personal safety, copyright preemption, civil conspiracy, and removal and alteration of CMI under the DMCA. Notably, the court found that while the plaintiffs had standing to seek injunctive relief because they sufficiently alleged that without an injunction there would be a

---

[1] A second class-action complaint was filed seven days later on behalf of two additional plaintiffs, J. Doe 3 and J. Doe 4, alleging similar claims against similar defendants. The two complaints were later consolidated. See _J. Doe 1 v. GitHub, Inc._, No. 22-cv-06823-JST (N.D. Cal. Jan. 3, 2024).

Motion To Dismiss Ruling Provides Insight Into How Courts View AI Training Data Cases

substantial risk that their code would be reproduced in Codex and Copilot's future outputs, the plaintiffs did not have standing to seek monetary damages because the developers had failed to demonstrate that Copilot's output reproduced their code. In other words, the plaintiffs did not allege that "they themselves have suffered the injury they describe[d]." Additionally, regarding the plaintiffs' claims under Section 1202(b) of the DMCA, which restricts the removal or alteration of CMI, the court found that the plaintiffs properly alleged that the defendants intentionally designed their programs to remove CMI from the output and that the plaintiffs raised a "reasonable inference that [d]efendants knew or had reasonable grounds to know that removal of CMI caried a substantial risk of inducing infringement." The defendants' motions to dismiss with respect to DMCA Section 1202(b) claims were denied in part and granted in part with leave to amend.[2]

First Amended Complaint

Following the court's ruling, the plaintiffs filed a first amended complaint, which generally tracked the original complaint with some modifications, including adding a fifth plaintiff, J. Doe 5, who similarly owned a copyright interest in code that was allegedly ingested, copied and reproduced by Copilot. To address the court's ruling on the defendants' motions to dismiss, the plaintiffs also added specific examples in which Copilot output the code of certain plaintiffs essentially verbatim or in a modified format "that contains only semantically insignificant variations" or a "modified copy that recreates the same algorithm." The defendants once again moved to dismiss portions of the plaintiffs' first amended complaint and the court issued its ruling on the defendants' motions in January 2024.

The Court's Ruling

As the court did with the defendants' first motions to dismiss, the court denied the defendants' motions to dismiss in part and granted them in part. The court reached the following key conclusions:

- **Regarding Article III standing violations:** The court found that by including examples in which Copilot generated as output code owned by Does 1, 2 and 5 in the first amended complaint, Does 1, 2 and 5 adequately alleged "particular personalized injury" to confer standing for monetary damages, so the court denied the defendants' motions to dismiss the plaintiffs' claims for damages due to lack of standing with respect to these three plaintiffs.[3] The court rejected the defendants' argument that a plaintiff's own actions (*i.e.*, inputting their own code to demonstrate output) cannot be used to demonstrate injury, stating that "a plaintiff is

not required to suffer an injury only inadvertently." The court also rejected the defendants' argument that the plaintiffs had failed to allege that users would seek to copy the developers' code. The court noted that for purposes of seeking standing for monetary damages, Article III does not require plaintiffs to demonstrate that users would want to reproduce their code or that users have entered or would be likely to enter the same prompts used in the plaintiffs' examples. The court suggested that these arguments more properly addressed the question of the amount of damages the plaintiffs had suffered.

- **Regarding DMCA Section 1202(b) infringements:** While the court previously denied the defendants' motions to dismiss plaintiffs' claims under Sections 1202(b)(1) and 1202(b)(3) of the DMCA, the defendants specifically asked the court to address an argument made in prior filings: "'[Section] 1202(b) claims lie only when CMI is removed or altered from an *identical* copy of a copyrighted work.'" The court agreed that Section 1202(b) has an identicality requirement and thus agreed with the defendants' argument that because the plaintiffs acknowledged that Copilot's output is more often a modification than a verbatim copy of the original code, the plaintiffs "effectively pleaded themselves out of their Section 1202(b)(1) and 1202(b)(3) claims." The court therefore granted the defendants' motions to dismiss the plaintiffs' claims under Sections 1202(b)(1) and 1202(b)(3) of the DMCA. While the court found it "unlikely that this deficiency could be cured," the court granted the plaintiffs leave to amend "out of abundance of caution."

- **Regarding other claims:** The court granted the defendants' motions to dismiss the plaintiffs' various state law claims, such as intentional and negligent interference with prospective economic relations, unjust enrichment, negligence and unfair competition on preemption grounds. While the plaintiffs tried to creatively recharacterize their claims, the court ultimately found that the core of the claims fell under the purview of the Copyright Act and did not include an "extra element" required to avoid preemption. The court dismissed these claims with prejudice.[4]

Takeaway Points

As the number of lawsuits against AI developers and platforms continues to increase, the court's ruling on the defendants' motions to dismiss offers further insight into how courts may address certain issues related to training AI models using copyrighted materials:

- **Actual examples of reproducing content are compelling.** The court was more persuaded, at least in the context of denying a motion to dismiss, by actual examples of the alleged reproduction of training data than by abstract arguments that this reproduction might occur.

---

[2] For more information on the district court's ruling, see our May 23, 2023, article.

[3] The court found that Doe 3 and Doe 4 failed to provide specific examples in which their code was output by Copilot; thus, they failed to allege standing for monetary damages. The court dismissed their request for monetary damages with prejudice, but these two plaintiffs still have standing to pursue claims for injunctive relief.

---

[4] The unfair competition claim was dismissed only to the extent that it was predicated on the plaintiffs' other state law claims.

Motion To Dismiss Ruling Provides Insight Into How Courts View AI Training Data Cases

- **Plaintiffs can generate examples of alleged infringement.** The court's decision that plaintiffs can seek standing for monetary damages even where they themselves entered the prompts to generate the allegedly infringing content could be significant given that this is the approach plaintiffs have used in other cases, including *Authors Guild v. OpenAI Inc.*, 1:23-cv-08292 (S.D.N.Y. 2023) (in which the plaintiffs entered prompts to generate allegedly infringing outlines of sequels to and/or derivatives of their works) and *The New York Times Company v. Microsoft Corporation*, 1:23-cv-11195 (S.D.N.Y. 2023) (in which the plaintiff entered prompts to recreate articles that had been published). In addition, plaintiffs in other training data cases will likely cite as precedent the court's holding that plaintiffs do not need to show, for standing purposes, that other users of an AI model would want to reproduce their content through that model. This option could be a key argument in cases where there is no evidence that the original content was otherwise commonly used.

- **CMI claims may require cases of identical reproductions.** For plaintiffs seeking to claim that CMI was removed or altered from their works when the works were reproduced, the court's decision provides important guidance that examples of an identical reproduction of the original work may be required before such a claim can be made.

As this case and others continue to progress, we expect to see further developments that provide greater clarity on the application of various laws, including intellectual property laws, to AI use.