

Ruling on Motion To Dismiss Sheds Light on Intellectual Property Issues in Artificial Intelligence

05 / 23 / 23

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

Summer associate **Ian Luo** contributed to this alert.

This memorandum is provided by Skadden, Arps, Slate, Meagher & Flom LLP and its affiliates for educational and informational purposes only and is not intended and should not be construed as legal advice. This memorandum is considered advertising under applicable state laws.

One Manhattan West
New York, NY 10001
212.735.3000

Many of the key intellectual property issues presented by artificial intelligence (AI), ranging from the use of copyrighted material as training data in AI-models to whether AI-generated works can be protected under copyright law, will likely only be resolved through court decisions, and possibly new legislation. The wheels of the judicial process will likely spin slowly in this nascent area, as they often do when new technologies enter mainstream usage.

However, a recent California district court decision on a motion to dismiss in *J. Doe 1 v. GitHub, Inc.* sheds some early light on how courts might approach certain of these issues, even though, because it was decided on a motion to dismiss, the court was simply focused on whether the plaintiffs adequately pled their various causes of action.

Background

Certain of today's AI models employ machine learning in which the functionality of the model is based on "studying" a large corpus of material called "training data." For models designed to generate computer code in response to a user's text prompts, the training data is comprised primarily of existing computer code. Two of these AI products are Copilot and Codex.

In November 2022, two developers filed a putative class action, using the pseudonyms J. Doe 1 and J. Doe 2, alleging that Copilot and Codex were trained on the plaintiffs' copyrighted computer code. The complaint names as defendants: GitHub, an open source platform owned by Microsoft on which the plaintiffs' code at issue was published, and which distributes Copilot; Microsoft as the owner of GitHub; and various OpenAI entities that programmed, trained and maintain Codex. According to the allegations in the complaint, Copilot requires Codex to function.

Since the plaintiffs' code was released under open source licenses, which generally do not restrict how the code can be used, the plaintiffs here could not assert that use of their code as training data was an infringing use — an argument that may be available to other copyright holders whose works are licensed under proprietary licenses and then used without permission as training data.

Instead, the plaintiffs' argued that 11 of the open source licenses a developer can opt to use on GitHub require that any derivative work or copy of the licensed work include attribution of the owner, and inclusion of a copyright notice and a copy of the open source license under which the code is licensed. Plaintiffs alleged that, when their code was used as training data, this information was stripped out. They also alleged that some of the AI-generated works by Codex and Copilot included portions of their copyrighted code.

The plaintiffs' complaint included a range of claims, including those for violations of the Digital Millennium Copyright Act (DMCA); violations of the GitHub terms of use; unfair competition; as well as claims that the plaintiffs' sensitive personal data was improperly used. The defendants moved to dismiss.

The Court's Decision

Were the Plaintiffs Injured?

A threshold issue was whether the plaintiff-developers suffered sufficient injury to satisfy Article III standing. The developers advanced two theories of injury: (1) that their personal information was sold and exposed (and would continue to be sold and exposed) by the defendants' actions and (2) that the use of their code as training data constituted harm to their property interests.

Ruling on Motion To Dismiss Sheds Light on Intellectual Property Issues in Artificial Intelligence

Privacy injury. The court quickly dismissed the privacy-based claim because the developers failed to identify the specific sensitive or private personal information at issue. Thus, the alleged facts were insufficient to show that the alleged misuse of the developers' personal data could give rise to a privacy injury. As a result, the court dismissed claims arising from the GitHub privacy policy, violations of the California Consumer Protection Act and negligence based on use of this information.

Property rights. The court devoted more attention to whether there was an injury to the plaintiffs' property rights. Here, the court focused on the issue that the injury alleged must be "particularized" (*i.e.*, that the plaintiff has itself suffered the injury in question), citing the Supreme Court's decision in *TransUnion LLC v. Ramirez*, 141 S. Ct. 2190 (2021). The plaintiffs asserted that their claim met this standard since they had alleged that in several instances Copilot's output matched licensed code written by a GitHub user. However, the court found this was an insufficient basis for injury since the plaintiffs failed to show that *their own code* had been included in that output.

The key takeaway from this part of the decision is the importance of drawing a direct connection between content that was allegedly used as training data and the output that was generated. Note that in *Anderson v. Stability AI, et al*, a case involving the use of various artists' works as training data, the defendants have moved to dismiss based on a similar argument.

Future harm. Interestingly, the plaintiffs also alleged that their allegations should survive a motion to dismiss based on the risk of "future" harm: *i.e.*, even if their works had not been included in Copilot output to date, it was likely to happen in the future.

The court acknowledged that the risk of future harm is a viable claim, but to allege monetary damages on future conduct, there must be an allegation of an additional, concrete harm, which the plaintiffs here failed to establish. However, the court agreed that the risk of future harm can be the basis for injunctive relief where the "risk of harm is sufficiently imminent and substantial" (citing *TransUnion*).

The court held that the plaintiffs had plausibly alleged that, without an injunction on Codex and Copilot's continued operations, there would be a substantial risk of those programs illegally reproducing the plaintiff's licensed code as output. This was based, in part, on allegations that GitHub's own internal research revealed that Copilot reproduced code from training data about 1% of the time, and that such output code did not reproduce license text, attribution and copyright notices, in violation of the open source licenses through which the plaintiffs licensed their code. The court thus allowed plaintiffs' claims to proceed based on future injury for which they were seeking injunctive relief.

Copyright Preemption

The defendants alleged that the plaintiffs' state law claims were all preempted by Section 301 of the Copyright Act, which preempts all state law claims that are within the subject matter of copyright and grant rights that are equivalent to the exclusive rights granted to copyright holders by the act.

Since most of the plaintiffs' state law claims were dismissed, the court focused on the preemption of the "unjust enrichment" claim. Plaintiffs maintained that their state law claims were qualitatively different because they also concerned "use" of their works (as training data), which is not a right granted by the Copyright Act. The court agreed with this theory, advanced in plaintiffs' opposition to the motion to dismiss, but noted that "use" was not actually alleged in the complaint. Rather, the complaint focused on reproduction and the preparation of derivative works, which are exclusive rights under the Copyright Act, and therefore preempted. The court dismissed the unjust enrichment claim with leave to amend.

The key takeaway here is that allegations of improper *use* of software should be able to survive a preemption challenge, at least in California.

Removal or Alteration of Copyright Management Information

Under the DMCA, the removal or alteration of copyright management information (CMI) is unlawful, as is distributing works knowing CMI has been removed if one has reasonable grounds to know it will induce infringement. CMI includes, as is the case in this matter, the identity of the copyright owner, the terms and conditions for use of a work, and other information that may be found in a copyright notice. (17 USC § 1202(b)).

The plaintiffs alleged that their code included CMI that the defendants removed or altered, and distributed despite having reasonable grounds to know that such actions would induce infringement. The defendants countered that "removal" of CMI requires an affirmative act and that the complaint merely alleged "passive non-inclusion of CMI." The court rejected this semantic distinction, and noted that the plaintiffs had properly alleged that the defendants were aware of the presence of CMI and had trained their programs to ignore it or remove it.

The defendants also argued that the developers had failed to sufficiently plead scienter (*i.e.*, that the defendants had knowledge that their actions would induce infringement). The court acknowledged that, although the "universal possibility" that an action might cause infringement is not sufficient, at the pleading stage, mental state does not need to be alleged with specificity.

Ruling on Motion To Dismiss Sheds Light on Intellectual Property Issues in Artificial Intelligence

Here, the court found that the plaintiffs had alleged that defendants knew the training data included CMI and knew that CMI was important to protect copyright interests. Thus, the court found that the plaintiffs' allegations raised a reasonable inference that the defendants knew or had reasonable grounds to know that removal of CMI carried a substantial risk of inducing infringement. The court therefore denied the defendants' motions to dismiss the §§1202(b)(1) and 1202(b)(3) claims relating to the removal or alteration of CMI.

The court did, however, grant defendants' motion to dismiss (with leave to amend) plaintiffs' claim that the defendants had distributed CMI knowing the CMI had been altered. (§1202(b)(2)) The court reviewed plaintiff's CMI allegations and found that it had failed to properly allege the distribution of altered CMI.

The key takeaway here is that removal or alteration of CMI, including for use in training data for an AI model, could potentially constitute a DMCA violation.

Breach of License

The defendants moved to dismiss the plaintiffs' claim that the use and distribution of their code in training data violated the open source licenses under which such code was licensed, arguing that the plaintiffs failed to allege with specificity which licenses were at issue or which provisions of those licenses had been breached as required under California law. The court denied defendants' motion to dismiss, finding that the plaintiffs had adequately set forth the 11 licenses that GitHub suggested for developers, and that these licenses included attribution requirements that defendants had breached when using the code as training data.

Unfair Competition

Plaintiffs' allegation of unfair competition was grounded in the Lanham Act and California statutory and common law, and predicated on violations of the DMCA, tortious interference, false designation of origin, violations of the CCPA, and negligence. Given that many of these predicate claims had already been dismissed, the court dismissed the corresponding unfair competition claim as well. Since, as noted, the court had not dismissed certain of the DMCA claims relating to removal of CMI, the court focused on whether this can form the predicate for an unfair competition claim.

The key question was whether the plaintiffs had properly pled that such violations also caused the plaintiffs' economic injury as required for an unfair competition claim. In their opposition to the motion to dismiss, the plaintiffs alleged a number of economic injury theories, including that they lost the value of their work; the likelihood they would be retained in the future

was impacted; and that they suffered injury to their intellectual property rights. The court did not determine the sufficiency of these injuries, but held that, since they were not alleged in the complaint and only raised in the plaintiffs' opposition to the motion to dismiss, the defendants' motion to dismiss would be granted with leave to amend.

Protecting the Pseudonymous Plaintiffs

The plaintiffs' use of pseudonyms in an intellectual property case is somewhat unusual, and defendants moved to dismiss based on the argument that plaintiffs cannot proceed under "John Doe" fictitious names. Plaintiffs responded that they had done so because of direct physical violence threats they had received through their counsel for pursuing this case. The defendants countered that the plaintiffs' fears were unfounded because the threats constituted simple, modern day internet trolling.

The court rejected this argument because the plaintiffs were subject to legitimate and credible threats of severe physical violence that would cause a reasonable person to fear harm, and the threats were directly and intimately targeted at the defendants, and were not mere provocative statements uttered in a public forum. The court also found that the defendants were not prejudiced at this stage of the litigation by plaintiffs proceeding pseudonymously, nor was there any harm to the public interest in allowing this.

Other Claims

The court denied defendants' motion to dismiss based on the contention that the plaintiffs had not pled sufficient facts regarding the role of each defendant in the alleged misconduct, finding that plaintiffs had done so.

The court did, however, dismiss plaintiffs' civil conspiracy claim with prejudice, because civil conspiracy is not a standalone cause of action, and only imposes liability on a defendant who did not itself commit a tortious act but agreed with third-party tortfeasors to partake in an illegal act.

Lastly, the court dismissed the developers' declaratory relief claim with prejudice because declaratory relief is also not an independent cause of action.

Final Thoughts

As noted above, there are some key takeaways from different aspects of the court's decision. More importantly, the decision provides a roadmap of what courts may expect to see at the pleading stage in cases involving the use of copyrighted materials as training data for AI models. As this and other cases proceed, the decisions at various stages will help shape the intersection between intellectual property law and AI.