AInsights

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

MacKinzie M. Neal
Associate / New York
212.735.2856
mackinzie.neal@skadden.com

One Manhattan West
New York, NY 10001
212.735.3000

Court Grants Motion To Dismiss in Kadrey AI Training Data Case

In a short but sharply worded decision, a California district court on November 20, 2023, granted the defendants' motion to dismiss in *Kadrey v. Meta Platforms, Inc*. The case is a putative class action brought by three authors — Sarah Silverman, Richard Kadrey and Christopher Golden — who alleged that Meta had used their copyrighted books to train LLaMA, a set of artificial intelligence (AI) large language models.

While the court granted the plaintiffs leave to amend most of their claims, its ruling all but shuts the door on some of the theories the plaintiffs had alleged.

Background

Meta had disclosed that the LLaMA training dataset included training data from a category called "Books," which came from two internet sources: (1) Project Gutenberg, an online archive of approximately 70,000 books that are out of copyright, and (2) "the Books3 section of ThePile … a publicly available dataset for training large language models."

While Meta did not further elaborate on the contents of those datasets, the plaintiffs alleged that other sources revealed that certain of their works were included in the Books3 dataset and therefore were part of the LLaMA training dataset.

Meta moved to dismiss all claims other than the one alleging that its act of copying the plaintiffs' books into the training set was itself direct infringement.

The Court's Ruling

The court reached the following key conclusions:

- **Claim that LLaMA is itself an infringing derivative work.** The court dismissed as "nonsensical" the plaintiffs' claims that the LLaMA models were themselves infringing derivative works because the "'models cannot function without the expressive information extracted from the plaintiffs' books." According to the court, there is no way to understand the LLaMA models themselves as a recasting or adaptation (*i.e.*, a derivate) of any of the plaintiffs' books.

- **Claim that all outputs are infringing derivative works.** The court similarly dismissed the plaintiffs' claims that every output of LLaMA is itself an infringing derivative work of the plaintiffs' works because these outputs are derived from such works, and that every output is an act of vicarious copyright infringement because users initiate queries of LLaMA to generate such outputs. The court noted that the complaint did not include

Court Grants Motion To Dismiss in Kadrey AI Training Data Case

an allegation regarding the content of any output that would support such a claim, and that without a plausible claim of infringing output, there cannot be vicarious infringement.

- The court rejected the plaintiffs' theory that because their books were copied to train LLaMA, they were not required to allege any similarity between the copied books and LLaMA outputs to sustain a claim of derivative infringement. The court held that a plaintiff is always required to establish that some portion of the original work is included in, or substantially similar to, the allegedly infringing derivative work.

- **Claims for DMCA violations.** The court dismissed the plaintiffs' Digital Millennium Copyright Act (DMCA) claims because they had failed to allege any facts that LLaMA distributed their books without their copyright management information, as required for a DMCA claim.

- **Other claims.** The court dismissed the plaintiffs' unfair competition law, unjust enrichment and negligence claims as preempted by their copyright law claims. The court also noted that to the extent the plaintiffs were seeking to survive preemption based on separate allegations of fraud or unfairness, they "have not come close to" alleging such conduct. The court also dismissed the plaintiffs' claim that Meta breached its duty of care "'to act in a reasonable manner toward others'" when it copied the plaintiffs' books to train LLaMA. The court not only found such a claim to be preempted but also cast significant doubt on whether such claim "could [even] be thought to exist."

Takeaways

Judge Vince Chhabria's decision in *Kadrey* closely follows the decision by Judge William H. Orrick, also in the U.S. District Court for the Northern District of California, largely dismissing similar claims by the plaintiffs in *Andersen v. Stability AI*. Both cases were filed by the same plaintiffs law firm.

Together, *Kadrey* and *Andersen* show that courts may have little appetite for copyright claims alleging that AI models are de facto derivative works of their underlying training data, or that AI-generated outputs are derivative works of such training data, without more direct factual allegations to establish these claims.

As noted in our November 2, 2023, client alert on the *Andersen* decision, other pending training data cases present factual allegations that address the shortcomings highlighted by the courts in *Kadrey* and *Andersen*.

Lawsuits by owners of copyright materials continue to be filed. On November 21, 2023, the author Julian Sancton filed a putative class action suit against OpenAI and Microsoft in the U.S. District Court for the Southern District of New York, alleging that his book had been included in the training set for OpenAI. In contrast to the *Andersen* and *Kadrey* suits that alleged a number of causes of action, Sancton's claims are limited to those alleging direct infringement and contributory infringement for copying a book he authored into OpenAI's training set.