# of artificial intelligence and untidy facts: federal district court denies summary judgment in ai copyright case

Robert H. Rotstein and Eleanor M. Lackman
MSK Client Alert

September 28, 2023

On September 25, 2023, Judge Stephanos Bibas of the U.S. Court of Appeals for the Third Circuit, sitting by designation in the District of Delaware, issued an opinion addressing the potential liability of an artificial intelligence startup for training its program using a copyrighted database. *Thomson Reuters Enterprise Centre Gmbh, et al. v. Ross Intelligence Inc*., Case no. 1:20-cv-613-SB (D. Del., Sept. 25, 2023). Noting that summary judgment is proper only when the facts are "tidy," the court concluded that the facts before it were too messy to avoid a jury determination. The decision marks an early judicial foray into the controversial issue whether an AI platform infringes a database used for training—and when suing copyrighted works for training constitutes fair use.

Plaintiff Thomson Reuters, which owns the Westlaw database, compiles judicial opinions according to Westlaw's "Key Number System" and adds headnotes that briefly summarize the relevant points of law that appear in the opinion. Plaintiff had a copyright registration for, among other material, the Key Number System and Headnotes. Defendant Ross sought to create a "natural language search engine" using machine learning and artificial intelligence. The search engine would avoid "human intermediated materials." As the court described Ross's system, "Users would ask questions and its search engine would spit out quotations from judicial opinions—no commentary necessary."

Of course, Ross needed material to train the machine. At first, it sought a license from Thomson Reuters to use Westlaw, but Plaintiff refused a license, unwilling to help another party create a completing platform. So, Ross turned to a third-party contractor to create memos with answers to legal questions that a lawyer would ask. This "Bulk Memo Project" resulted in approximately 25,000 question-and-answer sets. The third-party contractor created the memos both manually and, for a time, with the help of a text-scraping bot. The contractor also sent Ross a list of 91 legal topics from Westlaw's Key Number System. Ross admits

attorneys

Eleanor M. Lackman
Robert H. Rotstein

practice areas

Artificial Intelligence (AI)
intellectual property

msk.com

los angeles
t 310.312.2000
f 310.312.3100

new york
t 212.509.3900
f 212.509.7239

washington, dc
t 202.355.7900
f 202.355.7899

that it "considered" these topics when creating its own set of 38 topics that were used in an experiment, but ultimately abandoned that project. Finally, the contractor sent Ross 500 judicial opinions, including Westlaw's headnotes, key numbers, and other annotations. Ross claimed it did nothing with these opinions.

Thomson Reuters sued, contending that the questions in Ross's Bulk Memo Project were nothing more than Westlaw headnotes with question marks at the end. Ross responded that the headnotes "influenced" the questions but that lawyers had ultimately drafted them instead of copying them. The parties brought a total of five motions and cross-motions for summary judgment, each addressed to discrete issues.

Copyright Infringement. Ross first argued that because Plaintiff had just one copyright registration comprising hundreds of thousands of headnotes, copying only a few thousand was not enough for infringement. The court rejected this argument, noting that a copyright in a compilation extends to the copyrightable pieces of that compilation. Because "[h]eadnotes are just short written works, authored by Thomson Reuters…, they could receive standalone, individual copyright protection." However, the court found a genuine issue of disputed fact on whether the headnotes follow the uncopyrightable judicial opinions so closely as to be unoriginal. That latter issue would therefore go to a jury.

Turning to the issue of copying, the court held that Ross had copied portions of the Westlaw headnotes, both because Ross had admitted some copying and because Westlaw had shown access and probative similarity to the database. However, Judge Bibas went on to hold that the issue of substantial similarity of protected expression was a question for the jury: He simply could not decide as a matter of law whether there were similarities in *copyrightable* expression, especially in light of conflicting expert testimony. The court went on to find issues of disputed facts on Plaintiff's additional, more technical theories of liability for copyright infringement.

Fair Use. The parties brought cross-motions for summary judgment on the fair use issue, determining that all four fair use factors must go to a jury.

As to the first factor, the court found that Ross's use was, as a matter of law, commercial. Refusing to "overread" *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 143 S. Ct. 1258 (2023), however, the court nonetheless considered the question of "transformative use" highly relevant. Among other arguments, Ross contended that it had engaged only in intermediate copying to reverse engineer, which a number of previous decisions have found to be fair use. The court refused to apply a rigid rule that intermediate copying is always transformative, instead holding: "It was transformative intermediate copying if Ross's AI only studied the language patterns in the headnotes to learn how to produce judicial opinion quotes. But if Thomson Reuters is right that Ross used the untransformed text of headnotes to get its AI to replicate and reproduce the creative drafting done by Westlaw's attorney-editors," then the prior reverse-engineering cases are inapposite. Thus, the issue of transformative use was a question for the jury.

On the second fair use factor, the nature of the copyrighted work, the court held that because headnotes are not at the core of intended copyright protection, this factor tended to weigh in favor of fair use. However, the court also found this issue to be a jury question because of the uncertainty as to the headnotes' originality. As to the third fair use factor—substantiality of the use—a disputed issue of fact existed because it was unclear how much

msk.com

los angeles
t 310.312.2000
f 310.312.3100

new york
t 212.509.3900
f 212.509.7239

washington, dc
t 202.355.7900
f 202.355.7899

Ross actually took of the copyrighted material.

Finally, on the fourth factor, harm to Thomson Reuters' potential market, the court rejected Plaintiff's argument that Ross was a direct competitor. Seizing on language from *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183 (2021), Judge Bibas gave great weight to conflicting evidence as to whether Ross's AI platform had a "public benefit."

State Law Claims and other defenses: The court ruled that one claim for interference with contract was preempted by the Copyright Act but that others survived and raised questions for the jury. Likewise, the court refused to grant summary judgment for Ross on its defenses—most notably that the Westlaw Headnotes were not protected by copyright.

The most notable aspect of the opinion is its refusal to hold that intermediate copying is inevitably transformative —an issue likely to arise in numerous other copyright infringement cases involving AI. In addition, the court's emphasis on public benefit in assessing the fourth fair use factor, market harm, is of particular note, as the factor ordinarily focuses on economics and not the public value of a derivative work.

msk.com

los angeles
t 310.312.2000
f 310.312.3100

new york
t 212.509.3900
f 212.509.7239

washington, dc
t 202.355.7900
f 202.355.7899