



February 13, 2025

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

Jordan Feirman

Counsel / New York
212.735.3067
jordan.feirman@skadden.com

Mana Ghaemmaghani

Associate / New York
212.735.2594
mana.ghaemmaghani@skadden.com

MacKinzie M. Neal

Associate / New York
212.735.2856
mackinzie.neal@skadden.com

Shannon N. Morgan

Associate / New York
212.735.3711
shannon.morgan@skadden.com

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

Court Reverses Itself in AI Training Data Case

In an unexpected development, the judge in *Thomson Reuters v. Ross Intelligence*¹ — a case concerning the use of copyrighted material to train an artificial intelligence (AI) model — reversed much of his 2023 decision denying summary judgment on most issues, concluding that many issues could actually be resolved in Thomson Reuters’ favor on motions instead of going to a jury. While the case did not concern generative AI (*i.e.*, models that generate new content), some of the court’s reasoning may prove persuasive for copyright infringement cases involving generative AI models as well, including with respect to the fair use doctrine.

Background

Ross is the developer of a legal research tool that produces judicial opinion language in response to a user’s questions. Ross sought to license Thomson Reuters’ Westlaw content to train Ross’s AI model. When Thomson Reuters refused, Ross purchased question and answer pairings (referred to as “Bulk Memos”) from a third party and used those Bulk Memos to train its AI model. The Bulk Memos were distilled from Westlaw headnotes (*i.e.*, summaries of the key legal points in an opinion) by the third party. Thomson Reuters sued for copyright infringement and Ross claimed, among other defenses, fair use.

The parties cross-moved for summary judgment, and in a September 2023 decision, Third Circuit Judge Stephanos Bibas, sitting by designation in the District of Delaware, largely denied the motions, holding that most of the issues presented, including those regarding whether the headnotes were copyrightable and whether the fair use defense applied, were issues of fact to be determined by a jury. As the case was proceeding to trial, however, Judge Bibas explained that he “studied the case materials more closely” and asked the parties to renew their summary judgment motions.

The Court’s Decision

Copyrightability of Thomson Reuters Headnotes

In his September 2023 decision, Judge Bibas held that whether the Thomson Reuters headnotes were original enough to be copyrightable was a question of fact for the jury. Specifically, the jury would need to decide whether the headnotes met the originality threshold under copyright law or simply overlapped with portions of uncopyrightable judicial decisions.

In his current decision, Judge Bibas acknowledged that he was wrong to have focused solely on this overlap given the minimal amount of creativity required to secure copyright protection. Given that low threshold, the court found that Thomson Reuters’ selection and arrangement of its headnotes were copyrightable. Moreover, the court found that individual headnotes written by Thomson Reuters satisfied the originality standard even

¹ D. Del. Feb. 11, 2025.

Court Reverses Itself in AI Training Data Case

if the headnotes copied parts of an opinion verbatim since the editorial discretion of which “fraction of the whole” to include was enough of a “creative spark” to be protected.

Did Ross Copy Thomson Reuters’ Headnotes?

The court went through the exercise of comparing each of the 2,830 Bulk Memos against both the Thomson Reuters headnotes and the full judicial opinions. The court concluded that 2,243 of the analyzed Bulk Memos were clearly copied from the Thomson Reuters headnotes and not the original judicial opinions, and found substantial similarity to those headnotes, such that “no reasonable jury could find otherwise.” The court therefore granted summary judgment on the issues of actual copying and substantial similarity with respect to these 2,243 headnotes, and left for trial a number of other headnotes where factual issues remained about whether the Bulk Memos were copied from the judicial opinions themselves or whether there was substantial similarity between the Bulk Memos and the headnotes.

Was Ross’ Use of the Thomson Reuters Headnotes Fair Use?

Judge Bibas further reassessed whether Ross’s actions, even if infringing, are protected as fair use. Whereas he previously concluded that there were genuine factual disputes concerning whether fair use applied, Judge Bibas now reversed that holding based on “new information and understanding” and rejected Ross’s fair use defense.

In so finding, the court assessed the four statutory fair use factors: (1) the purpose and character of the use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the copying, and (4) the effect of the use upon the copyrighted work’s value and potential market. While the court acknowledged that fair use is a mixed question of law and fact, it found that “[t]he undisputed facts here push this case squarely into the legal realm.”

The court found that Ross’s use was “commercial” since it stood to profit from Thomson Reuters’ work, and was not transformative, at first glance, because Ross was creating a legal research tool to compete with Westlaw.

The court further considered whether Ross’s use of the headnotes could be considered transformative because Ross copied the headnotes only as an intermediate step in order to build its AI model. Specifically, Ross converted the headnotes into numerical data so that its model could recognize the relationships between

legal words. The actual text of the headnotes never appears in the AI model output since that output consists of language from judicial decisions.

While the court acknowledged that certain courts have previously found that intermediate copying supported a fair use defense, the court found those cases distinguishable because they concerned computer code that serves a functional purpose, as opposed to text, which does not. The court also noted that in cases where intermediate copying was allowed, such copying was necessary to innovate (*e.g.*, to discover the functional requirements for compatibility), which was not the case with Ross’s use of the Thomson Reuters headnotes.

Finally, the court found that the “impact on the market” factor favored Thomson Reuters since Ross’s use of the headnotes impacted Thomson Reuters’ business of legal research as well as its right to sell its data for training purposes. In its 2023 opinion, the court had left this to the jury to decide (“whether the public’s interest is better served by protecting a creator or a copier”). However, the court now reversed its stance since Ross had not put forth enough evidence that it was not developing a market substitute or that there was not a potential market for licensing out training data. With respect to the public interest in protecting the copier, the court concluded that copyright encourages people to create works like legal research tools, and Ross could have created its own legal search product without infringing Thomson Reuters’ rights.²

Key Points

It is important to note that the *Thomson Reuters* case was filed in 2020 — well before the explosion of generative AI models — and the functionality of Ross’s model is different from the more recent, high-profile cases involving AI training data. Most importantly, the Ross AI model is not a generative AI model in that it does not generate text in response to user queries; rather, its output consists of the judicial decisions themselves. Therefore, the application of the court’s decision to generative AI cases is likely limited. As Judge Bibas himself states: “I note for readers that only non-generative AI is before me today.”

However, there are two key points in the court’s decision that will likely be cited by plaintiffs in training data cases, and that could prove influential in forthcoming jurisprudence:

² The court found that the “nature of the work” factor favored Ross given the minimal creativity in the Thomson Reuters headnotes, as did the “amount and substantiality” factor since the headnotes themselves do not appear in Ross’s outputs. However, these factors did not sway the court’s ultimate conclusion.

Court Reverses Itself in AI Training Data Case

- The question of whether an algorithmic representation of text can be infringing has been raised in a number of the pending AI training data cases. Judge Bibas is now the second judge to find that the transformation of text into an algorithm when training an AI model does not necessitate a finding of non-infringement or fair use. Last August, in *Sarah Andersen v. Stability AI Ltd.*,³ Judge Orrick of the Northern District of California held that the fact that the plaintiffs' works may be included in an AI model in a different medium (*i.e.*, an algorithm) is not an impediment to the direct copyright infringement claim.

- Plaintiffs in AI training data cases are also likely to cite the growing market for licensing such data to support a claim that unauthorized copying has an effect on the market for purposes of a fair use analysis. Judge Bibas was receptive to such an argument, and took into account the effect on Thomson Reuters to license its own data for these purposes.

While the *Thomson Reuters* case presented certain unique facts, it is one more building block in the growing jurisprudence of AI training data cases.

³ *Andersen v. Stability AI Ltd.*, No. 23-CV-00201-WHO, 2024 WL 3823234 (N.D. Cal. Aug. 12, 2024)