AInsights

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

Pramode Chiruvolu
Counsel / Palo Alto
650.470.4569
pramode.chiruvolu@skadden.com

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

Brianna M. van Kan
Associate / New York
212.735.2177
brianna.vankan@skadden.com

Taylor Beardall
Law Clerk / Palo Alto
650.470.3101
taylor.beardall@skadden.com

One Manhattan West
New York, NY 10001
212.735.3000

525 University Avenue
Palo Alto, CA 94301
650.470.4500

In AI Training Case Brought by Thomson Reuters, Court Denies Summary Judgment

In *Thomson Reuters v. Ross Intelligence*,[1] a district court largely denied the parties' cross motions for summary judgment and held that a number of factual issues must be decided by a jury.

- The decision provides some insight into how courts might address the complex issue of whether the use of copyrighted works to train artificial intelligence systems constitutes copyright infringement and, if so, the viability of a fair use defense.

- Filed in 2020, long before the recent spate of similar cases,[2] *Thomson Reuters* is likely to be the first to go to trial on a critical legal issue in the development of AI systems.

- The main works at issue — Westlaw headnotes that summarize points of law from public domain court decisions — enjoy only narrow copyright protection.

- While the case may have limited precedential value for other "training data" cases where the works at issue enjoy broader copyright protection, it may still frame future discussions and company practices.

- The case also exemplifies how content owners may rely on their contractual rights when bringing claims concerning the unauthorized use of data for AI training purposes.

Background

Defendant Ross Intelligence, Inc. sought to develop a legal research tool that would output judicial opinion language in response to natural language queries.

Ross hired a third party, LegalEase Solutions, to create memos with model question and answer pairings (referred to as Bulk Memos) to train the AI engine driving Ross' platform. LegalEase, in turn, created the Bulk Memos by using the headnotes that Thomson Reuters provides with published court opinions on its Westlaw platform. Headnotes are summaries of the key legal points in an opinion, and can either be a direct quote of the opinion or include some Westlaw original expression. LegalEase held a license to use Westlaw.

---

[1] (D. Del. Sep. 25, 2023). Judge Stephanos Bibas was sitting by designation.

[2] Please see our client alerts "Ruling on Motion To Dismiss Sheds Light on Intellectual Property Issues in Artificial Intelligence" and "What Is Generative AI and How Does It Work?" for insights into some of these other cases.

In AI Training Case Brought by Thomson Reuters, Court Denies Summary Judgment

Thomson Reuters filed suit against Ross, alleging that the use of its headnotes to produce the Bulk Memos and then to train an AI model on those Bulk Memos infringed its copyright in the headnotes. Thomson Reuters also alleged Ross tortiously interfered with Thomson Reuters' contractual relationship with LegalEase by inducing LegalEase to breach the Westlaw licensing terms through automated text-scraping and password sharing.

Ross countered that its use of the headnotes is a fair use and that federal copyright law preempts the tortious interference claims.

Both parties moved for summary judgment.

Factual Issues Exist as to Whether Ross Engaged in Copyright Infringement

The court first addressed whether Ross had infringed Thomson Reuters' copyright in its headnotes, focusing on the three elements of such a claim: ownership of a valid copyright, actual copying and substantial similarity.

- **Ownership of a valid copyright.** Ross argued that since Thomson Reuters only holds a single compilation copyright in its vast array of headnotes, Ross' copying of a few thousand is insufficient to establish infringement. The court rejected that argument since the compilation registration extends to the individual headnotes as well. However, the court denied summary judgment and left it to a jury to decide:

  - The factual issue of whether the headnotes in question are protectable expression created by Thomson Reuters or simply close copies of the uncopyrightable judicial texts they annotate.

  - If Thomson Reuters' method of organizing and arranging judicial opinions through its Key Number System was protectable expression.

- **Actual copying.** The court ruled as a matter of law that Thomson Reuters had established actual copying not only through direct evidence (LegalEase admitted the copying) but also through circumstantial evidence since LegalEase had access to the headnotes and there was at least a "probative" degree of similarity between the Bulk Memos and the headnotes.

- **Substantial similarity.** The court declined, however, to grant summary judgment regarding whether there was substantial similarity between the Bulk Memos and the headnotes because factual disputes remained as to the extent of the protectable expression in Westlaw's headnotes and the substantiality of the similarity of the Bulk Memos to the headnotes, particularly since they "share an underlying source": uncopyrightable judicial opinions.

In sum, while the court found that Ross had engaged in actual copying, it left to a jury to decide the critical elements of the validity of Thomson Reuters' copyright and whether there was substantial similarity.

Thomson Reuters Copyright Infringement Claims Will Proceed to Trial

The court denied summary judgment on Thomson Reuters' direct, vicarious and contributory infringement claims, highlighting many of the same factual issues noted above.

- **Direct infringement.** With respect to direct infringement of Thomson Reuters' reproduction right, the court noted that the key facts were uncontested: Ross hosted copies of the Bulk Memos on its servers, copied them into Ross' machine learning "portal" and processed and labeled them. However, an open factual question remains as to whether the Thomson Reuters headnotes are protectable expression.

- **Vicarious and contributory infringement.** The court denied summary judgment on Thomson Reuters theories of contributory and vicarious liability because there remained factual disputes as to whether Ross knew LegalEase was infringing the headnotes or merely knew that it was using Westlaw, and the extent to which Ross was able to supervise and control LegalEase's conduct.

Disputed Facts Are Key to Fair Use Defense

The parties cross-moved for summary judgment on whether Ross' actions, even if infringing, were protected as fair use. The court denied both motions, running through each of the four fair use factors.

Purpose and Character of Use

With respect to the purpose and character of Ross' use of Westlaw materials, the court examined commerciality (which weighs against fair use) and transformativeness (which weighs in favor).

Commerciality

Thomson Reuters relied on the Supreme Court's recent decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*[3] to argue that Ross' commercial purpose of competing with Thomson Reuters weighed heavily against fair use. As noted in our client alert on *Warhol*, "Supreme Court Addresses Copyright Fair Use Defense in *Goldsmith*," we expect plaintiffs to routinely make this argument in fair use cases.

---

[3] 143 S. Ct. 1258 (2023).

In AI Training Case Brought by Thomson Reuters, Court Denies Summary Judgment

Transformativeness

Here, the court declined to "overread" *Warhol*, particularly since the *Warhol* court had recognized that transformativeness can outweigh the commercial character of the use. Instead, the court chose to focus on transformativeness and rely on the Supreme Court's holding in *Google v. Oracle*,[4] which it deemed more analogous to this technological context.

Ross argued that it transformed the underlying Westlaw material beyond recognition by converting the Bulk Memos into machine readable data, processing that data through an algorithm that trains its AI model about legal language, and producing a system that will return answers not only to the allegedly infringing questions, but to other legal questions a user might pose.

Ross further relied on cases holding that "intermediate copying," under which copying of protected material in order to discover unprotected information, or as a minor step in developing a whole new product, has previously been deemed a transformative fair use.

Thomson Reuters asserted that there was no transformation here because Ross had created a product to synthesize the law, no different from what Westlaw does, and that the translation of its headnotes into machine readable numerical data is a "paradigmatic derivative work."

Thomson Reuters also argued that the intermediate copying cases were inapposite since the defendants there were studying functionality or creating a compatible product, which was not the case here.

In what is likely the most significant portion of its decision, the court held that Ross would have engaged in transformative fair use if its AI merely studied language patterns in the Westlaw headnotes to learn how to produce judicial quotes in response to user questions, and not to replicate the Westlaw headnote themselves. While the court left it to a jury to make this factual determination, the legal framework established by the court weighs heavily in favor of finding fair use on this factor.

Nature of Copied Work

The second prong of the fair use analysis examines whether the nature of the copied work aligns closely with the core of copyright law's intended purposes. The court noted that this prong depends on the threshold open questions discussed above regarding the strength and extent of Thomson Reuters' copyright in the Westlaw headnotes. Although the court acknowledged that the expression in these headnotes is unlikely to be considered at the core of copyright protection, it left this factual question for a jury to decide.

Amount and Substantiality of Copying

With respect to the amount and substantiality of the copying, the court noted that the key inquiry was whether the copying took the "heart" of the original work since even a small amount of copying will not constitute fair use if the core piece of expression was copied. The court also indicated that this factor is closely linked to whether the use was transformative. For example, even verbatim intermediate copying has been found to be fair use where that copy is not revealed to the public.

The court held that this factor will come down to the factual issues of how Ross' AI works and the output it produces (*i.e.*, whether it is reproducing protected Westlaw expression or merely portions of the original unprotected court opinions). The court left this to a jury to decide. Notably, the court also held that Ross must demonstrate at trial that the scale of its copying "was practically necessary and furthered its transformative goals."

Effect on Value and Potential Market of Copied Work

Finally, the court reviewed the effect of using Westlaw materials on the value and potential market for those materials. While the court acknowledged that Ross' goal was to compete with Westlaw, the court noted that any market harm must be limited to the effect on Thomson Reuters' copyrightable expression, a factual question for a jury. The court also indicated that transformativeness is a key factor since the more an original has been transformed, the less likely there will be market impact.

Thomson Reuters argued that there was an impact on three markets here:

- The market for the Westlaw service itself (which the Ross product might supplant).
- The "traditional" market for licensing Thomson Reuters data (since Ross obtained Westlaw content through LegalEase).
- The potential market for Thomson Reuters licensing out its data for AI training purposes.

---

[4] 141 S. Ct. 1183 (2021).

In AI Training Case Brought by Thomson Reuters, Court Denies Summary Judgment

Ross argued that its platform was transformative in that it serves a different purpose than Westlaw, and that Westlaw would never participate in the market to license its data. The court once again held that a jury must resolve these factual disputes.

Notably, the court also highlighted questions regarding the public benefit of allowing fair use of copyrighted materials as training data. On one hand, it observed that permitting uses that result in a platform like Ross', which could enable access to the law at a lower cost, may provide a public benefit. On the other, it expressed concern that entities like Thomson Reuters could lose the incentive to create data like headnotes in the first place. This, too, the court viewed as a question more appropriate for a jury to decide as part of evaluating this fair use factor.

Tortious Interference Claims Partially Dismissed

In addition to its copyright infringement claims, Thomson Reuters asserted that Ross induced LegalEase to breach the Westlaw licensing terms restricting development of competing products, prohibiting use of text-scraping bots and sharing passwords.

Given the open factual disputes regarding these tortious interference claims, including regarding Ross' knowledge of the Westlaw terms and its role in causing breach of that contract, the court denied summary judgment and allowed these claims to proceed to trial.

Key Points

- Fair use cases are highly fact-dependent, and it is therefore not surprising that the court held that there were a number of factual issues that needed to be decided by a jury. However, the court's framing of Ross' fair use defense suggests that the court was persuaded that the unauthorized use of third-party content to train an AI model is a fair use where the model is not designed merely to replicate the original copyrighted content. Trial is tentatively scheduled for May 2024.

- A key factor is the narrow copyright protection Thomson Reuters enjoys in its headnotes. This contrasts sharply with the other AI training data cases brought to date, where the subject matter, such as books and images, is at the core of copyright. Nonetheless, early decisions — especially in areas of new technology — often frame the discussion and, sometimes, the practices of companies at the intersection of law and technology. Motions to dismiss have been filed by the defendants in many of the other cases, which will also shape the legal landscape.

- While much of the focus on the unauthorized use of data for AI training purposes has been on copyright infringement claims and the fair use defense, content owners are also relying on their contractual rights to prohibit this activity. The fact that the court did not find preemption of certain of Thomson Reuters' tortious interference claims indicates that use of bots and other techniques to collect training data in violation of contracts — whether that data is protected by copyright or not — may carry a degree of risk.