Thomson Reuters v. ROSS Intelligence (Summary Judgement)

In a closely watched decision revising a previous summary judgment, Judge Stephanos Bibas, a Third Circuit judge sitting by designation, sided largely with Thomson Reuters in its copyright dispute against ROSS Intelligence. The ruling granted partial summary judgment on direct copyright infringement claims while dismissing ROSS's argument that its use of Thomson Reuters' content qualified as fair use.

With Ross Intelligence now bankrupt and the technology at issue a decidedly niche application, attention is shifting to the broader implications for AI training and the use of copyrighted materials—particularly in the realm of generative AI. Earlier, Judge Bibas had refused to grant summary judgment on fair use, insisting the matter be put before a jury. However, upon further reflection, he reversed course, ultimately rejecting the defendant's fair use defense outright.

Background

Thomson Reuters, the owner of Westlaw, accused the AI-driven legal research firm ROSS of copyright infringement, alleging that it had improperly used legal summaries—so-called Bulk Memos—derived from Westlaw's editorial materials, particularly its headnotes, to train its technology. Thomson Reuters had refused to license its content to ROSS, a rival developing an AI-powered legal research tool requiring a database of legal questions and answers for training. To obtain the necessary data, ROSS partnered with LegalEase, which compiled and sold approximately 25,000 Bulk Memos—summaries created by lawyers referencing Westlaw headnotes. Whether the Bulk Memos involved verbatim copying or otherwise infringing copying was an issue in the case that ultimately went against ROSS. Upon discovering that ROSS had used content derived from these headnotes, Thomson Reuters filed a copyright infringement lawsuit. The summary judgment pertains only to a subset of the contested headnotes, leaving broader legal questions unresolved.

The court ruled against ROSS, determining that it had copied 2,243 headnotes and dismissing its various legal defenses, including claims of innocent infringement, copyright misuse, and the merger doctrine.

Ross's use was not transformative

Judge Bibas ruled that ROSS's use of Thomson Reuters' material was commercial and non-transformative, a conclusion that weighed heavily in the publisher's favor. According to the court, the use did not qualify as transformative because it lacked a distinct purpose or character from Thomson Reuters' original work.

The court's conclusion that Ross's use was not transformative is puzzling, especially given its acknowledgment—while discussing the third fair use factor—that the output of Ross's system did not replicate Westlaw's copyrighted headnotes but rather produced uncopyrighted judicial opinions.

The court did distinguish two significant cases, *Sega Enterprises Ltd. v. Accolade, Inc.* and *Sony Computer Entertainment, Inc. v. Connectix Corp.* but failed to consider cases like *iParadigms*, *HathiTrust* and *Google Books*. Even the way the court dealt with the reverse engineering cases is a bit suspect. The court sets them aside for two reasons, first because those cases involved copying software code, and second, that such copying was "necessary for competitors to innovate." To be sure, *Oracle v. Google* suggests that cases involving software may merit special treatment, but it is not clear why the software context should make a difference here. Judge Bibas's invocation of necessity is undercooked as well. Whether an act of copying is "necessary" is inextricably tied to the level of generality at which you ask the question. In *Oracle v. Google*, Google's replication of Java APIs was essential for compatibility with existing Java programmers, but whether that compatibility was a necessity or luxury depends on the level of generality at which you pose the question. After all, other smartphones ran without making life easy for Java programmers.

Not generative AI, but why?

The judge took care to distinguish this case from generative AI, yet the distinction remains murky. The court stated: *"Ross was using Thomson Reuters's headnotes as AI data to create a legal research tool to compete with Westlaw. It is undisputed that Ross's AI is not generative AI (AI that writes new content itself)."* And later that *"Because the AI landscape is changing rapidly, I note for readers that only non-generative AI is before me today."*

But what, exactly, sets this apart from generative AI? More broadly, how does this differ from other cases where nonexpressive uses have been deemed fair use? The opinion offers little guidance. It fails to engage with seemingly comparable precedents, such as plagiarism detection tools, library digitization for text analysis and digital humanities research, or the creation of a book search engine—cases where courts have found fair use.

The closest we get to an explanation of why Ross's use of the Westlaw headnotes is different to the intermediate copying *iParadigms*, *HathiTrust* and *Google Books* is that Ross merely retrieves and presents judicial opinions in response to user queries. This process, the court observed, closely parallels Westlaw's own practice of using headnotes and key numbers to identify relevant cases. Consequently, the court concluded that Ross's use was not transformative, as it primarily served to facilitate the development of a competing legal research tool rather than to add new expression or meaning to the copied material.

Market effect

The court determined that ROSS's actions impaired Thomson Reuters' market for legal AI training data, and in its reasoning, the fourth fair use factor carried substantial weight. Without qualification, the opinion echoes *Harper & Row*'s assertion that the fourth factor "is undoubtedly the single most important element of fair use." This is problematic. Asserting the absolute primacy of the Fourth factor is obviously in error in light of *Campbell*, as well as the Court's more recent decisions in *Google v. Oracle* and *Andy Warhol Foundation*. The Court's contemporary approach to fair use eschews rigid hierarchies among the statutory factors.

That said, the judge's finding in relation to the fourth factor may not be entirely unreasonable in this case: Ross explicitly intended to compete with Westlaw by creating a viable market alternative. For the court the key fact was that Ross "meant to compete with Westlaw by developing a market substitute." "And it does not matter whether Thomson Reuters has used the data to train its own legal search tools; the effect on a *potential* market for AI training data is enough."

Implications

One district court opinion that barely engages with the relevant caselaw will not change U.S. fair use law overnight, but it will certainly be welcome news for the plaintiffs in the more than 30 ongoing AI copyright cases currently being litigated.

I think what is really going on in this decision is that the judge has confused the first factor with the fourth factor. There is no obvious way to distinguish training on the question and answer memos to develop a model that directly links user questions to the relevant case law from cases involving search engines and plagiarism detection software. The real distinction, if there is one, is that ROSS used Westlaw's product to create a directly competing product.

Looking at the case this way, the decision might actually be good for the generative AI defendants, in cases like NYT v OpenAI, because there isn't the same direct competition.

* This is my first quick take on the decision just hours after it was handed down.

* Citation: Thomson Reuters Enter. Ctr. GmbH v. ROSS Intelligence Inc., No. 1:20-cv-613-SB (D. Del. Feb. 11, 2025)