IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

THOMSON REUTERS ENTERPRISE CENTRE GMBH and WEST PUBLISHING CORP.,

Plaintiffs,

v.

No. 1:20-cv-613-SB

ROSS INTELLIGENCE INC.,

Defendant.

Jack B. Blumenfeld, Michael J. Flynn, MORRIS, NICHOLS, ARSHT & TUNNELL LLP, Wilmington, Delaware; Dale M. Cendali, Eric A. Loverro, Joshua L. Simmons, KIRKLAND & ELLIS LLP, New York, New York.

Counsel for Plaintiffs

David E. Moore, Bindu A. Palapura, Andrew L. Brown, POTTER ANDERSON & CORROON LLP, Wilmington, Delaware; Gabriel M. Ramsey, Warrington Parker, Joachim B. Steinberg, Jacob Canter, Christopher J. Banks, Shira Liu, Margaux Poueymirou, Anna Z. Saber, CROWELL & MORING LLP, San Francisco, California; Mark A. Klapow, Lisa Kimmel, Crinesha B. Berry, CROWELL & MORING LLP, Washington, D.C.

Counsel for Defendant

MEMORANDUM OPINION

September 25, 2023

BIBAS, *Circuit Judge*, sitting by designation.

Facts can be messy even when parties wish they were not. But summary judgment is proper only if factual messes have been tidied. Courts cannot clean them up.

Thomson Reuters, a media company, owns a well-known legal research platform, Westlaw. It alleges that Ross, an artificial intelligence startup, illegally copied important content from Westlaw. Thomson Reuters thus seeks to recover from Ross. Both sides move for summary judgment on a variety of claims and defenses. But many of the critical facts in this case remain genuinely disputed. So I largely deny Thomson Reuters's and Ross's motions for summary judgment.

I. BACKGROUND

Many facts are disputed, but the basic story is not. Thomson Reuters's Westlaw platform compiles judicial opinions according to its Key Number System. That system organizes opinions by the type of law. Westlaw also adds "headnotes": short summaries of points of law that appear in the opinion. Each headnote is tied to a key number. Clicking on the headnote takes the user to the corresponding passage in the opinion. Clicking on the key number takes the user to a list of cases that make the same legal point. Westlaw has a registered copyright on its "original and revised text and compilation of legal material," which includes its headnotes and Key Number System. D.I. 255-7, at 8.

Ross Intelligence is a legal-research industry upstart. It sought to create a "natural language search engine" using machine learning and artificial intelligence. D.I. 310, at 4. It wanted to "avoid human intermediated materials." *Id.* Users would enter

questions and its search engine would spit out quotations from judicial opinions—no commentary necessary.

To leverage machine learning, Ross needed legal material to train the machine. At first, it tried to get a license to use Westlaw, but Thomson Reuters does not let users use Westlaw to develop a competing platform. So Ross turned to a third-party legal-research company, LegalEase Solutions. (LegalEase, in turn, hired a subcontractor, Morae Global. But the parties do not distinguish between LegalEase's and Morae's conduct, so I will refer only to LegalEase.)

Ross told LegalEase to create memos with legal questions and answers. The questions were meant to be those "that a lawyer would ask," and the answers were direct quotations from legal opinions. D.I. 310, at 4. The so-called Bulk Memo Project produced about 25,000 question-and-answer sets. Each memo had one question plus four to six answers and rated each answer's relevance. LegalEase created the memos both manually and, for a time, with the help of a text-scraping bot.

Ross says it converted the LegalEase memos into usable machine-learning training data. That involved first encoding the written language as numerical data and then running the data through a "Featurizer" that "performed various mathematical … calculations on the text." D.I. 272, at 8.

The core of this suit stems from the Bulk Memo Project. Thomson Reuters says the questions were essentially headnotes with question marks at the end. Ross admits that the headnotes "influence[d]" the questions but says lawyers ultimately drafted them, instead of copying them. D.I. 272, at 4–5. Though Thomson Reuters

contends that all 25,000 are copies, it has moved for summary judgment on just 2,830. It says LegalEase's copying of those 2,830 is undisputed because Ross's own expert admitted it.

Beyond the Bulk Memo Project, LegalEase provided Ross with two other relevant services. First, LegalEase sent Ross a list of 91 legal topics from Westlaw's Key Number System. Ross admits that it "considered" these topics when creating its own set of 38 topics that were used in an experimental "Classifier Project." D.I. 272, at 9–10. But it ultimately abandoned the Project. LegalEase also sent Ross 500 judicial opinions, including Westlaw's headnotes, key numbers, and other annotations. Ross says it did nothing with these opinions.

In this opinion, I address five summary-judgment motions. Thomson Reuters has moved for summary judgment on its copyright-infringement claim (limited to the 2,830 memos mentioned), and both sides have moved for summary judgment on Ross's fair-use defense. Thomson Reuters has also moved for summary judgment on its tortious-interference-with-contract claim, and Ross has counter-moved on its preemption defense to that claim.

"The court shall grant summary judgment if the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law." Fed. R. Civ. P. 56(a). A dispute is "genuine" if a reasonable jury could resolve it in favor of either side. *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 250 (1986). And a fact is "material" if it "could affect the outcome." *Lamont v. New Jersey*,

637 F.3d 177, 181 (3d Cir. 2011). I view the facts in the light most favorable to the nonmovant. *Id.* at 179 n.1.

II. COPYRIGHT INFRINGEMENT

A copyright-infringement claim has three elements: ownership of a valid copyright, actual copying, and substantial similarity. *See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 361 (1991); *Dam Things from Denmark v. Russ Berrie & Co.*, 290 F.3d 548, 561–62 (3d Cir. 2002). Here, all three elements are at least partly disputed. But the dispute over the second element is legal, so I can decide it now. And because Ross hired LegalEase to do the copying (if there was any), Thomson Reuters also couches its argument in terms of direct, contributory, and vicarious liability. So after addressing the three infringement elements, I will consider each of these liability theories as well.

A. The parties still dispute breadth and validity of Westlaw's copyright

Ross bets a good chunk of its infringement defense on Westlaw's being registered as a compilation. Ross's theory is this: because Westlaw has just one copyright registration, comprising hundreds of thousands of headnotes and key numbers, copying a mere few thousand is not enough for infringement.

Ross's gamble does not pay off. A copyright in a compilation extends to the copyrightable pieces of that compilation. *Educ. Testing Servs. v. Katzman*, 793 F.2d 533, 538–39 (3d Cir. 1986) (abrogated on other grounds) ("The fact that a registrant denominates the material as a compilation does not in itself signify that the constituent material is not also covered by the copyright."). And when the author of a compilation presents facts through his own original words, "[o]thers may copy the underlying facts

from the publication, but not the precise words used to present them." *Feist*, 499 U.S. at 348. Plus, though a plaintiff must have a registration to bring a federal suit for infringement, it can sue on all protected components of that one registration. 2 David Nimmer, *Nimmer on Copyright* § 7.16(B)(5)(c) (2023).

The cases Ross cites are the exceptions that prove the rule. In those cases, the copyright holder owned *only* the compilation. In one case cited, the plaintiff had a compilation copyright in the organization and selection of state legal forms. *Ross, Brovins & Oehmke, P.C. v. Lexis Nexis Grp.*, 463 F.3d 478, 480 (6th Cir. 2006). Though the underlying entries were in the public domain, the organizer's exact selection and arrangement were copyrightable. *Id.* Even though the defendant copied and compiled 61% of the forms from the plaintiff's compilation, there was no infringement because the defendant's compilation was not the "same selection." *Id.* at 483. So in these cases, the plaintiffs owned "thin" copyrights: other than their selection and arrangement choices, none of their compilations' components were protectable.

Here, only the Key Number System aligns with the compilation caselaw: It is Westlaw's method of organizing and arranging judicial opinions. So Thomson Reuters could have a valid copyright in this method of arrangement but not in the underlying opinions. That said, to qualify for copyright protection, "the manner of rearranging" and organizing the unprotectable underlying works "must constitute more than a minimal contribution." Nimmer, *supra*, § 3.03(A). This "threshold for originality is low," but the parties dispute facts needed to figure out if the System clears the bar. *Id.* § 3.04(B)(2)(a).

Thomson Reuters alleges that employees make creative organizing decisions to update and maintain the System and that the System is unique among its competition. But Ross replies that the System is unoriginal because most of the organization decisions are made by a rote computer program and the high-level topics largely track "common doctrinal topics taught as law school courses." D.I. 310, at 3. And although Thomson Reuters's registered copyright could protect its Key Number System, the jury needs to decide its originality, whether it is in fact protected, and how far that protection extends.

By contrast, the headnotes are not aptly described by the compilation caselaw. Headnotes are just short written works, authored by Thomson Reuters, so they could receive standalone, individual copyright protection. *See* 17 U.S.C. § 103. This distinguishes Thomson Reuters's copyright in its headnotes from the "thin," compilation-only copyrights in Ross's examples. So I must consider the alleged headnote copyright infringement at the level of each individual headnote, rather than at the level of the entire Westlaw compilation.

That said, Thomson Reuters's allegedly original expression in its headnotes still reflects uncopyrightable judicial opinions. So the strength of its copyright depends on how much the headnotes overlap with the opinions. Closely hewing differs from copying: If a headnote merely copies a judicial opinion, it is uncopyrightable. But if it varies more than "trivial[ly]," then Westlaw owns a valid copyright. *L. Batlin & Son, Inc. v. Snyder*, 536 F.2d 486, 490 (2d Cir. 1976).

The parties dispute how Thomson Reuters develops its headnotes and how closely those headnotes resemble uncopyrightable opinions. Thomson Reuters points to evidence that its headnotes are original representations of its attorney-editors' views—summarizing the most important case facts, highlighting key issues, and describing the holdings**.** Ross, though, presents evidence that Thomson Reuters's protocols required headnotes to "follow or closely mirror the language of judicial opinions." D.I. 272, at 13**.** This leaves a genuine factual dispute about how original the headnotes are. And this fact will serve double duty: it affects the strength and extent of Thomson Reuters's copyright, and it also goes to whether Ross was copying the headnotes or the opinions themselves.

In sum, I cannot decide the first element of Thomson Reuters's copyright infringement claim at summary judgment.

B. As a matter of law, Ross actually copied at least portions of the Bulk Memos

Next, Thomson Reuters must show that Ross (or LegalEase) "actually copied" its copyrighted work. "Actual copying focuses on whether the defendant did, in fact, use the copyrighted work in creating his own." *Tanksley v. Daniels*, 902 F.3d 165, 173 (3d Cir. 2018). If Ross "truly created [its] work independently, then no infringement has occurred, irrespective of similarity." *Id.* There are two ways to show actual copying: Thomson Reuters can present direct evidence. Or it can present circumstantial evidence demonstrating that Ross or LegalEase had access to the copyrighted work and that their work contains similarities probative of copying. *See id.*

Thomson Reuters presents both. LegalEase admitted to copying at least portions of the headnotes directly. As for circumstantial evidence, Ross does not dispute that LegalEase had access to Westlaw, which included access to headnotes. Though the similarities between Thomson Reuters's and Ross's work might not be substantial (that is a jury question), no reasonable jury could say that the similarities are not at least probative of some copying. And while Ross argues that any copying that occurred was miniscule in the grand scheme of the compilation, that framing misses the mark for the reasons given above. So Thomson Reuters has satisfied the actual-copying element as a matter of law.

C. Substantial similarity must go to the jury

The last element of direct infringement is substantial similarity. Substantial similarity asks whether "the ordinary observer, unless he set out to detect the disparities [in the two works], would be disposed to overlook them, and regard their aesthetic appeal as the same." *Id.* at 174 (alteration in original) (quoting *Peter Pan Fabrics, Inc. v. Martin Weiner Corp.*, 274 F.2d 487, 489 (2d Cir. 1960) (Hand, J.)). In other words, I ask whether an ordinary person would view the two works as basically the same.

This case features several wrinkles in the substantial-similarity analysis. First, the Bulk Memos could appear similar to Thomson Reuters's headnotes because they share an underlying source: uncopyrightable judicial opinions. But I must determine whether Ross's work is substantially similar to Thomson Reuters's *protected* expression, not just the opinions. Second, we contextualize the ordinary-observer test. *See id.* at 172 n.3. And here, the ordinary consumers of both parties' products are lawyers.

So I should be attuned to differences a lawyer might notice that a layperson might not. Finally, the Third Circuit "has [generally] rejected the usefulness of experts in answering" the substantial-similarity question. *Id.* at 172. I thus do not give much weight to the parties' dueling expert reports on this issue.

Substantial similarity "is usually an extremely close question of fact, which is why … summary judgment has traditionally been disfavored in copyright litigation." *Id.* at 171 (internal quotation marks omitted). Thomson Reuters argues that it can overcome this presumption because Ross's expert allegedly made an "admission." But this so-called admission does not get Thomson Reuters over the summary-judgment line.

The "admission" goes as follows: Ross's expert compiled the 25,000-plus Bulk Memo questions. She then paired each question with the Westlaw headnote most like it and paired each headnote with the judicial opinion passage most like the headnote. Next, on a scale of one to five, she rated two things: the similarity between the question and the headnote and the similarity between the headnote and the judicial opinion passage. The "admission" Thomson Reuters refers to is the 2,830 instances in which a question was rated as a close match to a post-1927 headnote, but that headnote was rated as not a verbatim or near-verbatim copy of judicial opinion text. (Copyrighted works created before 1927 are in the public domain and are not protected.)

Ross disagrees, breaking the headnotes into three groups. First, it says Thomson Reuters did not identify 1,623 of these 2,830 headnotes in its supplemental response, so they are not part of the case. Second, and more substantively, Ross says 1,019 questions are nearly identical to judicial opinions. Finally, it says nothing about

the remaining 188 entries, other than that they make up a tiny fraction of Westlaw's compilation. I take each group in turn.

First, I agree that Thomson Reuters did not identify the 1,623 headnotes. In an earlier Order, I reminded Thomson Reuters of its burden to show infringement and told it to identify "what, precisely, was copied." D.I. 201, at 1. Its response identified thousands of allegedly copied headnotes. But these 1,623 headnotes were not among those specifically identified. Thomson Reuters argues that it identified the Bulk Memos in which these headnotes were copied and that they incorporate by reference the cases in which the headnotes appear. But that is not precise enough. Producing the cases with the allegedly copied headnotes was what prompted Ross's objection and my Order to be more specific in the first place. So Thomson Reuters is limited (for purposes of summary judgment) to the 1,207 headnotes specifically identified.

There are genuine factual disputes over the second group. In her report, the Ross expert said each of these 1,019 questions overlapped a great deal with a headnote and that the headnote was not identical to opinion text. But she did not—and could not—take a position on whether the headnotes and questions were "substantially similar" under the ordinary-observer test. And more specifically, the report does not pinpoint how much similarity came solely from Thomson Reuters's protected expression.

Plus, Ross offers contrary evidence for these 1,019 entries. It shows that either the judicial opinion text is identical to the headnote or that the opinion text is more similar to the Bulk Memo question than the headnote is to the question. This supports the contention that similarity between Ross's and Thomson Reuters's work

stems from uncopyrightable judicial opinions, rather than from Thomson Reuters's original expression.

Thomson Reuters objects that Ross did not disclose its expert's methodology. But substantial similarity is not especially scientific: the question boils down to "good eyes and common sense." *Petrella v. Metro-Goldwyn-Mayer, Inc.*, 572 U.S. 663, 684 (2014). This makes Ross's mode of argument valid. Because both sides have lobbed conflicting expert reports (which I give little weight anyways) at each other and because a reasonable jury could agree with either side, I must send the question of substantial similarity for these 1,019 entries to the jury, where it typically belongs.

Finally, Ross does not object to the remaining 188 entries. Though substantial similarity is usually a close factual question, I will not review each of the 188 entries and make arguments for Ross. And for the reasons above, its reliance on Westlaw's copyright being solely in a compilation is misplaced. So each of these headnotes is substantially similar to its associated Bulk Memo question. But as noted above, whether this copying constitutes infringement depends on whether these headnotes are protected expression. And that rests on factual determinations the jury still must make. Plus, to recover from Ross, Thomson Reuters must win on one of its theories of liability and defeat Ross's fair-use defense.

D. All of Thomson Reuters's theories of infringement liability must go to trial

1. Direct liability. Thomson Reuters's theory of Ross's direct liability is uncontested: Ross hosted copies of the Bulk Memos on its servers, copied the content into its machine-learning "portal," transmitted another copy to a different server, created

more copies on employees' computers, then processed and labeled them by copying parts into another document. D.I. 250, at 13. Simply hosting a copy on a server might not seem like copying, but it is. *See MAI Sys. Corp. v. Peak Comput. Inc.*, 991 F.2d 511 (9th Cir. 1993); Nimmer, *supra*, § 8.08(A).

The unstated premise of this theory is that Ross violated Westlaw's reproduction right by making copies of the Bulk Memos. So for Thomson Reuters to succeed on direct liability, LegalEase's Bulk Memos must be unauthorized copies of protected expression. For making a copy of a non-copy is not copyright infringement. But because whether the Bulk Memos copied protected expression depends on factual determinations the jury must make, I cannot resolve direct liability at summary judgment.

2. Contributory liability. For Ross to be contributorily liable, Thomson Reuters must show that Ross (1) knew LegalEase was infringing and (2) materially contributed to or induced that infringement. *See Leonard v. Stemtech Int'l Inc.*, 834 F.3d 376, 387 (3d Cir. 2016). The parties dispute both prongs.

At best, Thomson Reuters has strong evidence that Ross knew LegalEase was using Westlaw. But knowledge or even encouragement to use Westlaw is not enough. One might expect a legal-research project to be completed using Westlaw, but merely using the service is not infringement. Plus, Ross points to evidence that it did not know LegalEase was infringing and never specifically instructed LegalEase to use Westlaw. Thomson Reuters has not done enough to prove that Ross knew about and

materially contributed to LegalEase's infringement. So this is not proper for summary judgment.

Thomson Reuters tries to bridge the gap between use and infringement by arguing that LegalEase breached its Westlaw license and Ross knew it. Once LegalEase breached the license, Thomson Reuters says, everything it did on Westlaw was copyright infringement. So Ross's knowledge of LegalEase's breach confers on Ross knowledge of the infringement. This argument mangles the interaction between licenses and copyright infringement.

In many copyright cases, licenses are used as a defense. In cases involving a license defense, one party claims infringement, and the other side claims they had permission through the license. But if the side claiming permission exceeded the scope of the license, it can be liable for infringement. *MacLean Assocs, Inc. v. Wm. M. Mercer-Meidinger-Hansen, Inc.*, 952 F.2d 769, 779 (3d Cir. 1991). That said, the copyright owner's rights must still be infringed by the specific activity that exceeds the scope of the license. Otherwise, the owner must litigate the license violations as breach-of-contract claims. *See S.O.S., Inc. v. Payday, Inc.*, 886 F.2d 1081, 1088–90 (9th Cir. 1989).

Here, there is no real dispute that LegalEase and Ross's alleged copying was not protected by a license. Rather, the issue is whether their actions constitute infringement of Thomson Reuters's copyright protections. And the license issue is irrelevant to proving Ross's contributory liability for LegalEase's infringement. So I deny summary judgment on the contributory-liability theory.

3. Vicarious liability. For vicarious liability, Thomson Reuters must show that Ross had "(1) the right and ability to supervise or control the infringing activity; and (2) a direct financial interest in such activities." *Leonard*, 834 F.3d at 388. Taking the elements in reverse, Ross does not contest that it had a financial interest in the alleged copies—it used the Bulk Memos to train AI, its core product. But it does contest whether it could supervise LegalEase. The control element is a matter of "practical ability." *Id.* at 389. So the determination is often fact-intensive. Evidence needs to support a finding that the defendant was "in a position to police the direct infringer[]." *Id.* at 388 (quoting *Fonovisa, Inc. v. Cherry Auction, Inc.*, 76 F.3d 259, 262–63 (9th Cir. 1996)). Thomson Reuters has testimony saying that Ross dictated LegalEase's practices. But Ross pushes back with evidence that LegalEase was secretive and resisted micromanagement. Thus, whether Ross had the "practical ability" to control LegalEase's "infringing activity" remains a disputed factual question for the jury to resolve.

III. FAIR USE MUST GO TO A JURY

The parties have cross-moved on Ross's fair-use defense. Fair use balances four factors: (1) the purpose and character of the use, (2) the nature of the copyrighted work, (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole, and (4) the effect of the use upon the potential market for the copyrighted work. 17 U.S.C. § 107. The first and fourth factors are most important. *See Authors Guild v. Google, Inc.*, 804 F.3d 202, 213–14 (3d Cir. 2015).

Fair use is a mixed question of law and fact. Though applying the test "primarily involves legal work," it requires "determination of subsidiary factual questions" about

the copying or the marketplace. *Google LLC v. Oracle Am., Inc.*, 141 S. Ct. 1183, 1199–200 (2021). Here, all of this must go to a jury.

A. The purpose and character of the use will be determined by contested facts

This first factor has two subparts: commerciality and transformativeness. *See Authors Guild*, 804 F.3d at 214, 218–19. (Bad faith is a minor subpart, also typically filed under this factor, and I will address it at the end.) Commercial use weighs against finding fair use, while transformative use weighs in favor. *Id.* at 218–20. And these considerations interact. "The more transformative the new work, the less will be the significance of … commercialism…." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994). Commerciality is straightforward: it asks whether the use was for profit. *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985). Transformativeness is less so: "a transformative use is one that communicates something new and different from the original or expands its utility, thus serving copyright's overall objective of contributing to public knowledge." *Authors Guild*, 804 F.3d at 214.

Ross's uses were undoubtedly commercial. And one of its goals was to compete with Westlaw. Thomson Reuters contends that this commercial use weighs heavily against finding fair use. In support of this claim, it cites the Supreme Court's recent decision in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 143 S. Ct. 1258 (2023). There, the Court determined that the use in question was not fair largely by emphasizing its commercial nature. *See id.* at 1279–80. But I decline to overread one decision, especially because the Court recognized that "use's transformativeness

may outweigh its commercial character" and that in *Warhol*, "both elements point[ed] in the same direction." *Id.* at 1280. Plus, just two terms ago, in a technological context much more like this one, the Court placed much more weight on transformation than commercialism. *Google*, 141 S. Ct. at 1204 ("[A] finding that copying was not commercial in nature tips the scales in favor of fair use. But the inverse is not necessarily true, as many common fair uses are indisputably commercial."). So I focus on transformativeness.

Thomson Reuters paints a black-and-white picture on transformativeness: Westlaw is a legal-research platform that synthesizes the law; Ross used Westlaw's syntheses to build a legal-research platform that also synthesizes the law. Ross, on the other hand, presents a more nuanced account: Westlaw headnotes and key numbers annotate opinions for users. Ross wanted to build a search engine that "avoids human intermediated materials," meaning a user would simply enter a query and get a responsive quotation from a judicial opinion, no clicking around or commentary needed. D.I. 272, at 1–3. Though Ross and Westlaw both help answer legal questions, Ross says it transformed the Westlaw headnotes beyond recognition.

Ross describes its process of transforming the Bulk Memos like this: First, it receives the Bulk Memos in its database. Then, it converts the plain-language entries into numerical data. Next, it feeds that data into its machine-learning algorithm to teach the artificial intelligence about legal language. The idea is that the artificial intelligence will be able to recognize patterns in the question-and-answer pairs. It

can then use those patterns to find answers not just to the exact questions fed into it, but to all sorts of legal questions users might ask.

Ross says that the caselaw on "intermediate copying" most appropriately reflects its use. In those cases, the users copied material to discover unprotectable information or as a minor step towards developing an entirely new product. So the final output—despite using copied material as an input—was transformative. In *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), the defendant copied Sega's copyrighted software. But it did so only to figure out the functional requirements to make games compatible with Sega's gaming console. *Id.* at 1522. That functional information was unprotected, so the copying was fair use. *Id.* at 1522–23.

Similarly, in *Sony Computer Entertainment Inc. v. Connectix Corp.*, 203 F.3d 596 (9th Cir. 2000), the defendant used a copy of Sony's software to reverse engineer it and create a new gaming platform on which users could play games designed for Sony's gaming system. *Id.* at 601. The court concluded that this was fair use for two reasons: the defendant created "a wholly new product, notwithstanding the similarity of uses and functions" between it and Sony's system, and the "final product [did] not itself contain infringing material." *Id.* at 606. The Supreme Court has cited these intermediate copying cases favorably, particularly in the context of "adapt[ing] the doctrine of fair use … in light of rapid technological change." *Google*, 141 S. Ct. at 1198 (cleaned up).

Thomson Reuters says the intermediate-copying cases are inapt. It argues that whereas in those cases, the copiers sought to "study functionality or create compatibility," here Ross simply sought to "train[] its AI" by "cop[ying] the creative decisions of West[law]'s attorney-editors precisely because it wanted to replicate them." D.I. 317, at 11. And it contends that Ross merely translated the headnotes into numerical data and that translation is "paradigmatic derivative work[]." D.I. 317, at 10.

But Ross says its AI studied the headnotes and opinion quotations only to analyze language patterns, not to replicate Westlaw's expression. So the translation was only a minor step in a broader, transformative use. *See Sega*, 977 F.2d at 1514–15, 1518–19 (holding that, though programmers wrote down and translated Sega's object code, these acts were a minor step towards a transformative use). If Ross's characterization of its activities is accurate, it translated human language into something understandable by a computer as a step in the process of trying to develop a "wholly new," albeit competing, product—a search tool that would produce highly relevant quotations from judicial opinions in response to natural language questions. This also means that Ross's final product would not contain or output infringing material. Under *Sega* and *Sony*, this is transformative intermediate copying.

So whether the intermediate-copying caselaw tells us that Ross's use was transformative depends on the precise nature of Ross's actions. It was transformative intermediate copying if Ross's AI only studied the language patterns in the headnotes to learn how to produce judicial opinion quotations. 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 Ross's comparisons to cases like *Sega* and *Sony* are not apt. Again, this is a material question of fact that the jury needs to decide.

Finally, the parties clash over whether Ross's use was in bad faith. But bad faith is at most a minor consideration in the fair use analysis. Indeed, the Supreme Court has expressed skepticism about whether it has any role to play at all. *Google*, 141 S. Ct. at 1204. And bad faith is particularly unimportant here. Thomson Reuters argues that Ross demonstrated bad faith by initially asking to license Westlaw, being denied, and then hiring LegalEase to illicitly gain access to it. But the Supreme Court has foreclosed this line of reasoning, explaining that "[i]f the use is otherwise fair, then no permission need be sought or granted. Thus, being denied permission to use a work does not weigh against a finding of fair use." *Campbell*, 510 U.S. at 585 n.18. So I can ignore bad faith. And the first fair use factor comes down to the jury's finding of transformativeness.

B. The nature of the copyrighted work favors fair use, but factual questions remain

The second factor asks about the nature of the copyrighted work. The work gets more protection, and copies are less likely to be fair, if it is near the "core of intended copyright protection." *Id.* at 586. But "[t]he scope of fair use is greater when 'informational' as opposed to more 'creative' works are involved." *Hustler Mag. Inc. v. Moral Majority Inc.*, 796 F.2d 1148, 1153–54 (9th Cir. 1986). So although judges should not act as critics, we consider "whether the work was creative, imaginative, and original." *MCA, Inc. v. Wilson*, 677 F.2d 180, 182 (2d Cir. 1981). That said, "[t]he second factor

has rarely played a significant role in the determination of a fair use dispute." *Authors Guild*, 804 F.3d at 220.

The analysis for this factor mirrors much of my earlier discussion of the validity and strength of Thomson Reuters's copyright. As explained above, this depends largely on factual questions that the jury must decide, so I cannot resolve this factor at summary judgment.

But I note here that the Key Number System is far from the core of copyright. Even if the system involves making creative decisions about how to organize opinions and other material and is an original method of organization, it is merely a way to arrange "informational" material. So the system inherently involves significantly less creative or original expression than traditionally protected materials, such as literary works or visual art, and is much less "imaginative."

The headnotes are closer, but still not especially close to the core. "The law generally recognizes a greater need to disseminate factual works than works of fiction or fantasy." *Harper*, 471 U.S. at 563. And though editors may have made creative choices about which points of law to summarize, how to summarize them, and where to attach the headnote, those choices are constrained. In general, the headnotes will flag the most salient points of law, largely track the language of the opinion, and be placed at the beginning of a paragraph. This approach is akin to news reporting, which, though protected, must be carefully separated from the unprotected underlying facts. *See Author's Guild*, 804 F.3d at 220. So, although a jury must decide how closely headnotes reflect the language of judicial opinions and, in turn, precisely how much

protection they are afforded, they are not at the core of intended copyright protection. Thus, although an ultimate decision on factor two must wait until trial, this factor seems to favor fair use.

C. The amount and substantiality of the copying depends on the nature of Ross's AI outputs

Third, I consider the amount of copying as well as whether the copying took the original work's "heart." *Campbell*, 510 U.S. at 589.

Defining the work at issue matters in determining the amount of copying done. If we define it at the level of each headnote, the copying was allegedly completed for some 25,000 headnotes. If we define it at the level of the compilation, however, the copying was less substantial, though headnotes likely represent the "heart" of Westlaw's expression.

And defining the use is again important because "even a small amount of copying may fall outside of the scope of fair use where the excerpt copied consists of the heart of the original work's creative expression." *Google*, 141 S. Ct. at 1205 (internal quotation marks omitted). Conversely, "copying a larger amount" can still be fair use "where the material copied captures little of the material's creative expression." *Id.* Plus, "[t]he 'substantiality' factor will generally weigh in favor of fair use where … the amount of copying was tethered to a valid, and transformative, purpose." *Id.* In particular, verbatim intermediate copying has consistently been upheld as fair use if the copy is "not reveal[ed] … to the public." *Authors Guild*, 804 F.3d at 221; *see also A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 638–640, 642 (4th Cir. 2009).

Here, the best definition is at the level of each headnote. As mentioned, the compilation registration also covers individually copyrightable materials. And each headnote counts. But the heart of each headnote is its original expression, not its link to the part of the opinion it summarizes. So if Ross's AI works the way that it says, it is likely fair use because it produces only the opinion, not the original expression. "It cannot be said that a revelation is 'substantial' in the sense intended by the statute's third factor if the revelation is in a form that communicates little of the sense of the original." *Authors Guild*, 804 F.3d at 223.

Yet this factor also requires jury fact-finding. How Ross's AI works and what output it produces remain disputed. The parties also fight over whether the use was "tethered to a valid … purpose." Westlaw says Ross copied far more than it needed. Ross says it needed a vast, diverse set of material to train its AI effectively. Though Ross need not prove that each headnote was strictly necessary, it must show that the scale of copying (if any) was practically necessary and furthered its transformative goals. So the third factor hinges on the answers to these disputed factual questions which the jury needs to resolve.

D. I cannot yet determine the effect of the use upon the market for the work

Finally, factor four asks whether the use had a "meaningful or significant effect" on the value of the original or its potential market. *Authors Guild*, 804 F.3d at 224. And "[t]his inquiry must take account not only of harm to the original but also of harm to the market for derivative works." *Harper*, 471 U.S. at 568. Yet not all losses are created equal. I must also consider the "source of the loss." *Google*, 141 S. Ct. at

1206. Again, we come back to the fundamental premise that copyright protects expression. If the source of the loss is not that the original's expression is being appropriated, "the type of loss of sale envisioned above will generally occur in relation to interests that are not protected by the copyright." *Authors Guild*, 804 F.3d at 224.

And transformativeness feeds into this factor as well. "[T]he more the copying is done to achieve a purpose that differs from the purpose of the original, the less likely it is that the copy will serve as a satisfactory substitute for the original." *Id.* at 223 (citing *Campbell*, 510 U.S. at 591). Finally, in evaluating market impact, courts must pay special attention to "the realities of how technological works are created and disseminated." *Google*, 141 S. Ct. at 1199.

Here, those "realities" are disputed. Thomson Reuters claims three potential markets, but they boil down to two: the market for Westlaw itself as a legal research platform and the market for its data. It says Ross's plan all along was to create a substitute for Westlaw. And it says that this plan worked, as some Ross customers cancelled their Westlaw subscriptions. As for the market for its data, Thomson Reuters says there is a traditional licensing market and a burgeoning one for AI training data. It argues that it lost traditional licensing revenue because Ross obtained Westlaw content through LegalEase. And it suggests that there is a potential market for Westlaw's training data; after all, Ross paid LegalEase over a million dollars for the Bulk Memos. That burgeoning market would be harmed by copying like Ross's.

One fact is undisputed here: Ross and Thomson Reuters both compete in the market for legal research platforms. But that alone does not reveal whether Ross's AI

product is a substitute for Westlaw. Ross's use might be transformative, creating a brand-new research platform that serves a different purpose than Westlaw. If so, it is not a market substitute. Ross also argues that Thomson Reuters has never participated—and would never participate—in this market for its training data. Because a reasonable jury could find for either side on these factual market-impact questions, I cannot resolve them at summary judgment.

Finally, "we must take into account the public benefits the copying will likely produce." *Google*, 141 S. Ct. at 1206. And "we are free to consider the public benefit resulting from a particular use notwithstanding the fact that the alleged infringer may gain commercially." *Sega*, 977 F.2d at 1523. This "[p]ublic benefit need not be direct or tangible, but may arise because the challenged use serves a public interest." *Id.*

The parties provide competing narratives of public benefit. Ross's research platform might increase access to the law at a lower cost. Or it might just reduce the incentives for Thomson Reuters, and similarly situated entities, to create content like headnotes in the future.

Deciding whether the public's interest is better served by protecting a creator or a copier is perilous, and an uncomfortable position for a court. Copyright tries to encourage creative expression by protecting both. Here, we run into a hotly debated question: Is it in the public benefit to allow AI to be trained with copyrighted material?

The value of any given AI is likely to be reflected in the traditional factors: How transformative is it? Can the public use it for free? Does it discourage other creators

by swallowing up their markets? So an independent evaluation of the benefits of AI is unlikely to be useful yet, even though both the potential benefits and risks are huge. Suffice it to say, each side presents a plausible and powerful account of the public benefit that would result from ruling for it. So a jury must decide the fourth factor—and the ultimate conclusion on fair use.

IV. TORTIOUS INTERFERENCE

Thomson Reuters's second claim is tortious interference with contract. It says Ross induced LegalEase to breach three contract provisions by (1) using Westlaw to build a competing product, (2) using a bot to scrape Westlaw content, and (3) sharing passwords.

Ross says these claims are preempted. Federal copyright law preempts "all legal … rights that are equivalent to any of the exclusive rights within the general scope of copyright as specified by section 106 … and come within the subject matter of copyright." 17 U.S.C. § 301(a). Section 106 protects the author's rights in reproduction, distribution for sale, public performance and display, and derivative works. *See id.* § 106.

Although preemption is an affirmative defense, I address it first. If any of the contract claims are preempted by copyright, I need not address their merits.

A. Thomson Reuters's first tortious-interference claim is preempted, but the other two survive

As quoted above, federal copyright law preempts state claims that are "equivalent to" § 106 rights. The most common test—and the one I must apply—to determine equivalency is the "extra element" test. *See Dun & Bradstreet Software Servs., Inc. v.*

Grace Consulting, 307 F.3d 197, 217–18 (3d Cir. 2002). As its name suggests, that test asks whether the state-law claim has an element that a § 106 claim would not. *Id.* at 217.

The test is easy to say but hard to apply. In some sense, every claim other than a state copyright claim has some extra element. For example, although almost every common law claim requires damages, § 106 does not because federal law supplies statutory damages. But that alone does not make the claims meaningfully different. So courts have modified the test, asking whether the extra element makes the claim qualitatively different. *Id.* To avoid preemption, the "gravam[e]n" of the state claim must differ from one of the § 106 rights. *Id.* at 218 (quoting *Computer Assocs. Int'l, Inc. v. Altai, Inc.*, 982 F.2d 693, 717 (2d Cir. 1992)).

Ross says tortious-interference claims are almost always preempted. Though it cites good authority for that argument, it takes that authority out of context. A tortious-interference claim that says something like, "You copied our work, thus interfering with our contracts to license that work" is preempted because it merely identifies one of the consequences of a § 106 violation. But some of Thomson Reuters's claims are different. It brings some of its claims as tortious-interference claims—rather than as standard breach-of-contracts claims—solely because it seeks to hold Ross liable for the acts of a third party, LegalEase. So the preemption analysis should look more like it would with a typical breach-of-contract claim than with a claim that just identifies a consequence of copyright infringement.

To sum up, a claim is preempted if (1) the material is within the subject matter of copyright and (2) the gravamen of the claim is equivalent to a § 106 right.

The anti-competition claim is preempted. Thomson Reuters's first tortious-interference claim concerns this provision:

> You may not sell, sublicense, distribute, display, store or transfer [West's] products or any data in [its] products in bulk or in any way that could be used to replace or substitute for [its] products in whole or in part or as a component of any material offered for sale, license or distribution to third parties.

D.I. 316, at 4 (alterations in original).

Thomson Reuters says Ross induced LegalEase to breach this provision by hiring it to create the Bulk Memos and other materials sent to Ross. Thomson Reuters does not contest that this covers material that is within the subject matter of copyright. But it says the rights implicated are different. Not so.

The gravamen of this claim is the same as that of Thomson Reuters's copyright claim. And the contract provision itself secures equivalent—indeed, sometimes identical—rights: The rights to sell, sublicense, distribute, and transfer are covered by § 106(3). The right to display is covered by § 106(5). The "in bulk" and "in any way that could be used to replace or substitute" phrases are incorporated in the fair-use analysis. And the "as a component of" language also tracks fair use and the derivative-work right under § 106(2).

Though this contract provision is framed in terms of competition, it is focused on one potential competitive threat: copying. That concern is the domain of federal law. So this first claim is preempted by the Copyright Act.

Thomson Reuters tries to analogize its provision to the ones at issue in cases like *Altera Corp. v. Clear Logic, Inc.*, 424 F.3d 1079 (9th Cir. 2005), and *Wellness Publishing v. Barefoot*, No. 02–3773, 2008 WL 108889 (D.N.J. Jan. 9, 2008). But those cases involved restrictions on use. Restricting a user's use of copyrighted material is different from limiting the user's ability to copy it. The latter is covered, and thus preempted, by the Copyright Act.

The anti-bot and password sharing claims are not preempted. The two other tortious-interference claims involve Westlaw's anti-bot and password-sharing provisions:

> You may not run or install any computer software or hardware on [West's] products or network or introduce any spyware, malware, viruses, Trojan horses, backdoors or other software exploits.
>
> Your access to certain products is password protected. You are responsible for assigning the passwords and maintaining password security. Sharing passwords is strictly prohibited.

D.I. 316, at 4–5 (alteration in original).

These provisions are not equivalent to § 106 rights. Unlike the competition provision, they govern use and manipulation of the site. Using a bot to scrape content might copy material in bulk. And a claim based on the harm from that copying itself would be preempted. But a claim based on simply introducing malware, independent of that malware's goals, is not equivalent to any right in § 106. Likewise, a site might ban password sharing because they want to limit copying risk. But putting limits on access to the site is a separate restriction. Whether the material behind the password protection is copyrighted or not, the creator can protect the material for which it

charges users. Section 106 has nothing to say about that limit. So Thomson Reuters's second and third tortious-interference claims survive preemption.

B. Thomson Reuters's two remaining tortious-interference claims are partially disputed

I now consider the two surviving tortious-interference claims on the merits. Thomson Reuters must prove five elements: (1) there was a contract between LegalEase and Westlaw, (2) Ross knew about the contract and its terms, (3) Ross's intentional act was a significant factor causing the breach, (4) Ross had no justification, and (5) the breach harmed Thomson Reuters. *See WaveDivision Holdings, LLC v. Highland Cap. Mgmt., L.P.*, 49 A.3d 1168, 1174 (Del. 2012). (Earlier in this case, the parties disputed whether Minnesota or Delaware law applied. But I need not address this choice-of-law question because Minnesota's tortious-interference elements are the same as Delaware's. So I address each element below under Delaware law.)

1. There was a contract. For both claims, there is no dispute that the first element is met—there was a contract between Westlaw and LegalEase. But three of the four remaining elements involve genuine factual disputes that remain for trial.

2. It is unclear how much Ross knew. To satisfy the second element, Ross must have had actual or imputed knowledge of the substance of the contract rights, even if it did not know about the exact terms. *WaveDivision*, 49 A.3d at 1176. There is substantial record evidence that Ross knew at least something about Westlaw's contracting practices: Ross itself tried to enter a contract with Westlaw. An investor sent Ross a copy of the Westlaw terms and conditions. A Ross executive posed as an individual

practitioner to see the terms of use. And there are email exchanges in which Ross executives discuss specific provisions of the contract.

But Ross's evidence introduces some ambiguity. It disputes the timeline, saying that much of Thomson Reuters's evidence comes from well after Ross dealt with LegalEase. And it says that although it saw the contracts Westlaw offered it, it never saw Westlaw's contract with LegalEase. Westlaw's agreements can be tailored, Ross says, and it only ever saw the Canadian, not United States, agreement. Thomson Reuters counters that the agreements are materially the same, publicly disclosed, and seldom if ever altered. Whether this evidence, taken together, rises to the level of knowledge of the substance of the anti-bot and password-sharing provisions is a jury question.

3. Ross may have intentionally caused a breach. Ross met the third element if it (1) intended to interfere or (2) intended to reach a result with knowledge that it would interfere with the contract, even if interference was not the main purpose. Restatement (Second) of Torts § 766 cmt. j (Am. L. Inst. 1965). Ross intended for LegalEase to produce the Bulk Memos. And it was likely aware that LegalEase was using Westlaw. But whether Ross knew LegalEase was breaching or going to breach by using a bot or sharing passwords is far less clear. As explained in the discussion of indirect liability, each side has evidence suggesting different levels of Ross's involvement, control, and knowledge. So this element too is unfit for summary judgment.

4. Whether Ross acted without justification is a factual question. Courts commonly refer to the fourth element as doing something "not … sanctioned by the rules of the

game." *Avaya Inc., RP v. Telecom Labs, Inc.*, 838 F.3d 354, 383 (3d Cir 2016) (internal quotations omitted). Thomson Reuters said Ross did that. According to Thomson Reuters, Ross first sought a license from Westlaw, but when that rule-abiding approach failed, it hired LegalEase to end-run Thomson Reuters's terms. Yet end-running or exploiting a loophole is not necessarily the same as violating the rules of the game. And whether Ross acted without justification largely depends on how it acted—which is a matter of dispute under the second and third elements. Plus, this element is typically a question of fact: Delaware law considers no less than seven factors, nearly all of which involve some underlying question of fact. *See WaveDivision*, 49 A.3d at 1174. So this element remains for trial.

5. Thomson Reuters has shown harm. The fifth element requires harm. Thomson Reuters says it lost subscription fees when LegalEase used a bot and shared passwords. If LegalEase did not have the help of a bot or multiple employees sharing one account, it would have had to buy more subscriptions or keep its subscriptions open for longer periods. This harm is distinct from the harm from copying: assuming copying was going to happen, Westlaw at least wanted to get paid while LegalEase did it. The bot and password sharing made the copying more efficient and cheaper, depriving Westlaw of fees. Ross does not contest this element, other than arguing that Thomson Reuters's damages here are the same as the damages it would get for its copyright infringement claim. That misses the mark, so there is no genuine dispute.

In sum, Thomson Reuters is entitled to partial summary judgment on the first and fifth elements of tortious interference by using a bot and sharing passwords, but elements two through four remain for trial.

V. Other Defenses Fail

Ross throws several other defenses at the wall, but none sticks. First, it no longer presses its First Amendment or first-sale defenses. Second, it raises laches. But laches does not apply to the copyright claim. *See Petrella*, 572 U.S. at 679. Plus, Thomson Reuters brought its tortious-interference claim promptly. Ross says that Thomson Reuters sat on its right to sue for years, but this is unsupported, and Ross cites no law saying that was too long.

Third, for its defenses of consent, waiver, estoppel, acquiescence, and license, Ross points to a fair-use provision in Westlaw's terms of use. That provision just begs the fair-use question but does not provide an independent defense. Fourth, Ross alleges tort of another, saying it was not a "substantial factor" in the harm that LegalEase allegedly caused. D.I. 318, at 20. But again that argument is the same as its argument against the elements of the tortious-interference claim. And it has presented no evidence specifically for this defense, so I will not repackage other evidence for it. Instead, Ross can focus on defeating the tortious-interference claim's elements directly.

Finally, Ross alleges lack of ownership because the headnotes are identical to public law. But Thomson Reuters has provided its registrations, and the extent of its expression is fully explored under the infringement and fair-use claims. Indeed, whether "lack of ownership" is even an affirmative defense is dubious—it seems to go to the first element of an infringement claim (ownership of a valid copyright). So I

grant summary judgment to Thomson Reuters on these miscellaneous affirmative defenses.

* * * * *

Thomson Reuters alleges that Ross copied protected aspects of Westlaw, both directly and indirectly through LegalEase. And Ross disputes almost all of Thomson Reuters's story. But it is not my role at summary judgment to sort through the evidence and tidy these factual messes. It is the jury's role at trial. So, with the small exceptions noted throughout this opinion, I deny both Ross's and Thomson Reuters's motions for summary judgment.