

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

THOMSON REUTERS ENTERPRISE
CENTRE GMBH and WEST PUBLISH-
ING CORP.,

Plaintiffs,

v.

No. 20-cv-613-SB

ROSS INTELLIGENCE INC.,

Defendant.

Jack B. Blumenfeld, Michael J. Flynn, MORRIS, NICHOLS, ARSHT & TUNNELL LLP, Wil-
mington, 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. Stein-
berg, Jacob Canter, Christopher J. Banks, Shira Liu, Margaux Poueymirou, Anna Z. Sa-
ber, CROWELL & MORING LLP, San Francisco, California; Mark A. Klapow, Crinesha
B. Berry, CROWELL & MORING LLP, Washington, D.C.

Counsel for Defendant

MEMORANDUM OPINION

September 28, 2023

BIBAS, *Circuit Judge*, sitting by designation.

Thomson Reuters and Ross filed six motions in limine. Each motion seeks to exclude expert testimony. So I consider them under Federal Rule of Evidence 702 and *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993). Expert testimony must come from a sufficiently qualified expert, use a reliable methodology, and reasonably fit the testimony and the case. *Elcock v. Kmart Corp.*, 233 F.3d 734, 741 (3d Cir. 2000). I consider each motion in turn to see if the expert testimony satisfies those requirements.

In summary: I did not consider expert testimony about damages or substantial similarity in my summary-judgment opinion. So, for the testimony about those subjects, I defer ruling on the motions until trial. I grant one part of the motion to exclude Alan Cox's testimony but deny all other motions.

A. Jonathan Krein's Testimony

Ross moves to exclude Jonathan Krein's testimony. I deny two parts of its motion. First, Ross argues that Krein improperly testified about an ultimate issue. Krein testified that LegalEase systematically copied Westlaw content and that Ross must have known about it. True, those assertions broach the legal issues of actual copying and contributory liability. But an expert may opine on an ultimate issue if his testimony will "assist the trier of fact to understand the evidence...." *In re Japanese Elec. Prods. Antitrust Litig.*, 723 F.2d 238, 278 (3d Cir. 1983), *abrogated on other grounds*; see FED. R. EVID. 704(a) ("An opinion is not objectionable just because it embraces an ultimate issue.").

Krein’s statements about copying can assist the factfinder. To determine the extent of copying, the factfinder may need to make sense of thousands of documents. Krein’s testimony could simplify that task; it draws conclusions from those documents. Ross seeks to exclude Krein’s conclusion that Ross systematically copied Westlaw content. But if Krein’s substantial-similarity methodology is reliable—a question I reserve for trial—then the summary of his results could be useful to the factfinder, even if it raises an ultimate issue. I will not exclude it just because it draws a conclusion about the breadth of LegalEase’s copying.

Plus, Krein’s testimony about Ross’s knowledge may help frame the relevant facts within the context of industry norms. The factfinder could credit Krein’s expertise in the technology industry. And if it does, it might rely on Krein’s testimony to understand how a company like Ross would typically supervise an important data-development subcontract. So Krein’s testimony that Ross must have known about LegalEase’s copying may help the factfinder “understand the evidence” about Ross’s knowledge and supervision of LegalEase’s actions. *In re Japanese Elec. Prods. Antitrust Litig.*, 723 F.2d at 238.

Second, Ross argues that Krein is not qualified to opine on the market for legal AI training data. It says Krein is an expert in technical software and computer science, not the AI data market. But the Third Circuit applies a liberal standard to assess an expert’s qualifications. An expert is qualified when he has “substantially more knowledge in the field” than the average layperson. *See Elcock*, 233 F.3d at 742–44. Krein studied machine learning and has professional experience in the technology

industry. So he can certainly outpace the average layperson's knowledge of the AI data marketplace. Thus, I deny this part of Ross's motion and leave its challenge to Krein's substantial-similarity testimony for trial.

B. James Malackowski's Testimony

I deny one part of Ross's motion to exclude James Malackowski's testimony. Like Krein, Malackowski stated that there is a potential marketplace for legal AI training data. Ross argues that Malackowski lacks any factual basis for this opinion, so it is unreliable. But Malackowski identified his factual sources: Krein's testimony about the AI training data market, documents reflecting Westlaw's and Ross's use of legal AI training data, and evidence of the developing marketplace for legal AI tools. Ross's quibble is really with the accuracy of the facts underlying Malackowski's opinion, not with whether the facts exist at all. Cross-examination can address that concern. So I deny that part of Ross's motion and hold in abeyance the remaining part related to Malackowski's damages testimony.

C. L. Karl Branting's Testimony

Thomson Reuters seeks to exclude L. Karl Branting's testimony on three grounds. All fail. *First*, it claims that Branting, in his deposition, expressly disavowed several opinions from his report. Branting's report contained statements about the contents of the Bulk Memos (the question-and-answer sets that Ross used as training data). But later, Branting testified that he did not know how the Bulk Memos were created and could not determine whether they contained copied Westlaw content. Still, Thomson Reuters's objection misses the mark because Branting's report focuses on the content of Ross's source code, not its training data.

Branting asserts that Ross’s code does not retain the text (the allegedly copied headnotes) or any external organization system (the allegedly copied Key Number System) from its training data. So, at its core, Branting’s opinion may help the factfinder understand whether Ross’s search engine mimics or reproduces Westlaw content. It does not provide new evidence about whether the Bulk Memos were copied in the first place.

Plus, the challenged statements in Branting’s report serve as factual inputs to his source code analysis, not as independent expert opinions. And he relies on facts that are supported by other evidence in the record. So, in his deposition, Branting did not disavow his whole expert report.

Second, Thomson Reuters argues that Branting lacked evidence to support his opinions about the content of the Bulk Memos. Like its disavowal argument, Thomson Reuters mischaracterizes the purpose of these statements. Branting relies on these statements merely as the factual underpinning of his core opinion: whether the Ross search engine reproduces Westlaw’s text or organization system. And “questions regarding the factual underpinnings of the expert witness’s opinion affect the weight and credibility of the witness’s assessment, not its admissibility.” *Bresler v. Wilmington Tr. Co.*, 855 F.3d 178, 195 (4th Cir. 2017) (cleaned up). So I deny this part of the motion.

Third, Thomson Reuters seeks to discredit Branting’s opinion that Ross’s program uses different AI features than Westlaw. It argues that Branting lacked a factual basis for the opinion—he did not know how Westlaw’s AI functions. But Branting

knew some information about Westlaw's AI. He supports his factual assertions by citing Thomson Reuters's operational documents and its employees' deposition testimony. Perhaps other facts—unknown to Thomson Reuters after discovery—would undermine his opinion. But that can be shown through cross-examination. Because Branting had a sufficient factual basis to testify about how Ross's and Westlaw's AI differed, I deny this part of the motion.

D. Barbara Frederiksen-Cross's Testimony

I hold in abeyance Thomson Reuters's motion to exclude Barbara Frederiksen-Cross's testimony. Thomson Reuters says Frederiksen-Cross used a reply report to improperly introduce new opinions and methodologies. I did not consider that reply report in my summary-judgment opinion, so I reserve judgment on this motion.

E. Richard Leiter's Testimony

I deny Thomson Reuters's motion to exclude Richard Leiter's testimony. Leiter describes the history of the Westlaw Key Number System. Thomson Reuters thinks his opinions are unreliable and irrelevant. But a historian's methodology can be reliable if it synthesizes historical sources to construct a "reliable narrative about th[e] past." *United States v. Kantengwa*, 781 F.3d 545, 562 (1st Cir. 2015) (internal quotation marks omitted). Leiter does that. Thomson Reuters may contest the historical evidence that he relies on, but that does not make his testimony inadmissibly unreliable.

Though Leiter's testimony might not get at the heart of the parties' factual dispute, it could help the factfinder understand the strength of Thomson Reuters's copyright. So there is a "connection between the expert opinion offered" and one of the

“particular disputed factual issues in the case.” *In re TMI Litig.*, 193 F.3d 613, 670 (3d Cir. 1999). That meets the relatively low threshold to establish “fit.” *See United States v. Ford*, 481 F.3d 215, 219 (3d Cir. 2007). Leiter’s testimony is both reliable and relevant enough to satisfy the *Daubert* inquiry.

F. Alan Cox’s Testimony

I deny one part of Thomson Reuters’s motion to exclude Alan Cox’s testimony, but I grant another part of it. I discuss the denial first. Thomson Reuters attacks Cox’s market benefit testimony. It says Cox’s opinion was too speculative. But Cox is an economist—he is qualified to opine about theoretical market dynamics. And although his analysis is terse, he applies well-established economic theory: a low-cost market entrant (i.e., Ross) could increase the total market for a good and segregate the market such that a high-quality seller (i.e., Thomson Reuters) can increase prices. So this opinion is reliable.

But Cox asserts another opinion that is not. He claims that a potential market for legal AI training data is highly unlikely. His methodology cannot support this conclusion.

To identify whether a market exists, Cox embraces a well-regarded methodology: First, define the product based on its attributes. Then, define the market for the product based on contextual factors like the existence of buyers and sellers, the cost of substitutes, and regulatory factors. But he makes an unreasonable assumption. He rejects a *potential* market because he cannot find a *current* one. And this flaw runs throughout his methodology.

For starters, Cox fails to define the product attributes of Westlaw Content—as it could be used in a potential market for legal AI training data. In his deposition, he admitted that he did not know how to define the AI training data. And to investigate whether buyers and sellers existed, he ran Internet searches for evidence that AI developers wanted to train their AI with Westlaw Content or something similar. Those searches cannot be replicated rigorously and reveal that Cox was imprecisely casting around for a product market.

Cox does suggest, however, that Westlaw Content has no market substitutes. That would alleviate the pressure on this analytical step. If Westlaw Content is a market unto itself, Cox could define the market without identifying its relevant product attributes because all attributes would be relevant. But he cites only Thomson Reuters's expert to assert that Westlaw Content has no substitutes. And that expert said only that a data set must be tailored to its purpose, not that each data set constitutes a unique market. At most, Cox shows that there are no current substitutes for Westlaw Content. But Cox has no factual basis to say that Westlaw Content has no potential, future substitutes. And if another expert tried to identify potential substitutes using Cox's methodology, she "would be lost" without a description of how Cox defined Westlaw Content's relevant product attributes. *See Elcock*, 233 F.3d at 747.

Then, Cox assumes that market conditions today foreclose a potential market. True, if Westlaw Content has no substitutes, a potential market might be foreclosed if Thomson Reuters refuses to sell to competitors. But there could be substitutes. So a potential market could emerge if another seller develops substitutable legal AI

training data. Of course, Cox could disagree with that assessment. But he did not provide a testable or replicable method to draw that conclusion. So I grant that part of Thomson Reuters's motion. I leave for trial the portions of the motion that seek to exclude Cox's disgorgement and lost profits opinions.

* * * * *

Each party complains about the other's experts. Many of the experts—but not all—satisfy Rule 702 and *Daubert*. So I deny most parts of the motions in limine. I reserve for trial several motions related to damages and substantial similarity. And I grant one part of Thomson Reuters's motion to exclude Alan Cox's testimony.