

**IN THE UNITED STATES DISTRICT
COURT FOR THE DISTRICT OF
MASSACHUSETTS**

**IN RE: ZELIS REPRICING ANTITRUST
LITIGATION,**

Lead Action Case No.: 1:25-CV-10734-BEM
Honorable Brian E. Murphy

This Document Relates To:

Consolidated with Case Nos.:

All Actions

1:25-CV-11092-BEM

1:25-CV-11167-BEM

1:25-CV-11537-BEM

**STIPULATED PROTOCOL GOVERNING PRODUCTION OF RESPONSIVE
INFORMATION USING TECHNOLOGY ASSISTED REVIEW (“TAR”)**

The Parties, by and through their undersigned counsel, agree that the protocol set forth herein shall govern their use of Technology Assisted Review (“TAR”) in connection with the production of documents in this action, should any Party choose to use TAR. No Party is required to use TAR in connection with its review and production of documents. Moreover, by agreeing to use TAR in this action and to follow this protocol, the Parties do not intend to waive any rights or protections pursuant to privacy, confidentiality, attorney-client privilege, attorney work product, and any other privileges, protections, or objections to discovery.

Nothing in this protocol is intended to or shall alter the duties of any Party under the Order Regarding Production of Electronically Stored Information and Paper Documents (Dkt. [XX]) (“ESI Protocol”), Stipulated HIPAA-Qualified Protective Order (Dkt. 253), Stipulated Preservation Order (Dkt. 255), or any other orders entered or to be entered by the Court.

I. Definitions

1. Technology Assisted Review” or “TAR” means the use of supervised machine learning algorithms to classify, prioritize, and categorize documents based on coding decisions made by human reviewers, encompassing TAR 1.0 (“Simple Active Learning” or “SAL”) and/or TAR 2.0 (also known as “Continuous Active Learning” or “CAL,” in which the machine learning model continuously retrains and re-ranks the document population as reviewers make coding decisions throughout the review). By way of example and without limitation, TAR 2.0 includes tools such as Relativity Active Learning, Brainspace Continuous Multimodal Learning (“CMML”), Everlaw Predictive Coding, and similar supervised machine learning features offered within litigation support platforms.

II. Documents Subject to TAR

2. The Producing Party may apply search terms to custodial sources and non-custodial sources that (i) are agreed to by the Parties or ordered by the Court (ii) for the agreed upon relevant discovery period. All documents and families that hit on the agreed upon search terms may be subject to TAR, should a Party elect to use it.

3. Should the Producing Party and Requesting Party be unable to reach agreement on the set of search terms used to determine the documents subject to TAR, the Parties shall promptly meet and confer, and if unable to resolve the dispute, either Party may submit the matter to the Magistrate Judge within 30 days of reaching impasse. Additionally, if documents produced in this litigation or other circumstances reveal that the search terms agreed upon by the Parties are failing to capture responsive materials, the Parties shall promptly meet and confer regarding supplemental or alternative search terms. Should the

Parties be unable to reach agreement following such meet and confer, either Party may submit the dispute to the Magistrate Judge within 30 days of reaching impasse.

4. After de-NISTing in accordance with Section IV.A of the ESI Protocol, the Producing Party will review the file types present in the document review population and provide the Receiving Party with a list of any additional file extensions to be excluded from TAR.

5. The Producing Party will run an analysis and based on aspects such as text size, the Producing Party will identify files that may not be suitable candidates for TAR, such as files with too little or too much extracted text, and files with other problematic extracted text), images, audio visual files, etc.

- a. Promptly following the analysis undertaken in Paragraph 5, the Producing Party will identify the files that are not suitable for TAR and meet and confer with the Requesting Party regarding the same.

6. The Producing Party will determine whether any documents can be successfully or adequately subjected to OCR processing and included in the TAR process. Documents that are not amenable to successful, adequate, or accurate OCR processing or which contain no substantive OCR text will be excluded from TAR review.

- a. Promptly following the determination in Paragraph 6, the Producing Party will identify the types of documents and file types excluded from TAR under this Paragraph.

7. The Producing Party will manually review documents excluded from TAR, including non-text based materials such as audio and video files, or otherwise notify the Parties if it believes any documents are not appropriate for TAR or manual review.

8. The Parties will meet and confer regarding any additional culling beyond what is contemplated here or already agreed upon by the Parties.

9. The Producing Party will disclose the total number of documents in the collection(s) and the total number of documents subject to TAR.

III. Prioritization and Training

10. The Parties will use document review attorneys familiar with this litigation to iteratively train a TAR model to identify potentially responsive ESI. Apart from the Control Set described in Paragraph 15 below (applicable to SAL (TAR 1.0), only), all responsive-based coding decisions will be used for training and incorporated into the TAR model. Any document that a Party, after a good-faith effort, identifies as coded incorrectly shall be fed back into the TAR system with the proper coding.

11. In the absence of an appropriate document selection algorithm within the TAR platform, the Producing Party will make a good faith effort to incorporate review and training on documents with diverse concepts in order to prevent bias in the predictive coding model.

12. The Producing Party will review documents prioritized by scoring/ranking, as continuously updated by training from reviewed documents either from an automatically updated queue or regularly generated, rank-limited batching based on current model ranking.

13. Quality control (“QC”) measures will be utilized regularly, including evaluating documents marked by the review team as responsive, yet were not ranked highly by the system or vice versa.

A. CAL (TAR 2.0)

14. The Producing Party using CAL (TAR 2.0) may train the predictive coding model with an initial document set. The Producing Party will review a statistically valid random

sample (95% confidence level, 5% confidence interval) of the TAR universe (“richness sample”).

B. SAL (TAR 1.0)

15. For a Producing Party using SAL (TAR 1.0), at the outset, a Subject Matter Expert (“SME”) will review and code a randomly selected set of 2,000 documents (“Control Set”) for responsiveness. SMEs will code the Control Set documents as responsive or non-responsive. The documents in the Control Set are not used to train the TAR model but serve as a benchmark to measure how well the model is classifying documents as responsive or non-responsive, through each training round by comparing how the model classifies documents against how the SME coded documents. The Control Set is used throughout the training process to assess the model’s predictions so training can be terminated when the benefits of additional training no longer outweigh the cost of additional training. The Control Set must have an adequate number of responsive and non-responsive examples to meet a statistical confidence level and margin of error.

16. The Producing Party using TAR 1.0 will review and code at least 1,750 documents randomly drawn from the TAR universe to serve as an initial seed set to train the TAR model. For clarity, the 1,750 randomly drawn documents used for the seed set must consist of a different random sample than that used for the Control Set.

17. Training the TAR 1.0 model will be an iterative process that takes place in rounds. After the initial seed set review is complete, the Producing Party will select additional document sets for SMEs to review using random and judgmental sampling techniques. The Producing Party using TAR 1.0 shall implement reasonable quality control measures between training rounds to gauge how training is progressing, including by comparing the

Control Set to model's predictions for responsiveness for those same documents, to test the model's accuracy.

18. Training will continue until the Producing Party concludes it has reached an 70% recall rate, with a minimum of 60% precision. The Producing Party will periodically sample to validate effectiveness of the model prior to finalizing the TAR 1.0 model.

19. If new custodians or ESI sources are added to the TAR universe following review of the initial seed set, the Producing Party using TAR 1.0 will review and code a random sample of 200 new documents for each new custodian or ESI source to update the initial seed set.

20. The Producing Party will provide the Receiving Party with statistics based on the final cut-off for recall (the percentage of total responsive documents retrieved by the model at a specific cut-off), precision (the percentage of documents retrieved confirmed as responsive), and margin of error. The review will not be complete if the validation efforts fail to demonstrate at least 70% recall or demonstrate a deficient process. If any of the above occur, the Producing Party will make a good faith effort to resolve the deficiency through means that may include selective document searches or additional training or refining of the TAR 1.0 model and repeating the validation process.

IV. Stopping Criteria

21. The review process should incorporate quality-control and quality-assurance procedures to ensure a reasonable production consistent with the requirements of Federal Rule of Civil Procedure 26(g). The TAR review process will continue until the Producing Party reasonably concludes that further review is unlikely to yield additional responsive documents with sufficient quantity or materiality to justify continuing. This will not occur

before the last three to four batches of documents, consisting in total of a minimum of 1,000 documents identified by TAR and reviewed by humans, contain no more than ten to fifteen percent (10%-15%) responsive documents. The TAR model shall be retrained at least three times and at regular intervals during the last three to four batches. Once a Producing Party reasonably believes that it has produced or identified for production substantially all responsive non-privileged documents, it shall conduct validation according to the sampling protocol described below and in Section V.

22. If the Producing Party believes further review is unlikely to yield additional responsive documents with sufficient quantity or materiality to justify continuing the review, the Producing Party shall notify the Requesting Party and promptly produce to the Requesting Party the responsive documents identified in the last two batches of review (the “Last Batches”) before stopping the review pursuant to Paragraph 20 above, and the recall estimation described in Paragraph 28 below, including Bates numbers of any responsive documents identified per the validation protocol described in Section V, below.

V. Validation

23. The review process shall incorporate quality-control and quality-assurance procedures to ensure productions comply with the requirements of Federal Rule of Civil Procedure 26(g). Once a producing Party reasonably believes that it has identified substantially all responsive documents, it shall conduct validation according to the sampling protocol described herein.

24. The Review Corpus is defined as including all documents collected for review and subject to TAR. The Review Corpus shall be portioned into the following Subcollections:

- a. **Subcollection C(1):** All documents subject to TAR identified by the review as relevant, including any privileged documents, but not including family members of

relevant documents, unless those family members are deemed to be relevant in their own right.

- b. **Subcollection C(2):** All documents subject to TAR coded as non-responsive by a human reviewer.
- c. **Subcollection C(3):** All documents subject to TAR excluded from manual review as the result of a TAR/ process.

25. A sample shall be drawn consistent with the following:

- a. **Subsample D(1)** shall consist of 400 documents randomly selected from Subcollection C(1).
- b. **Subsample D(2)** shall consist of 400 documents randomly selected from Subcollection C(2).
- c. **Subsample D(3)** shall consist of 1,200 documents randomly selected from Subcollection C(3).

26. Should a producing Party believe that the sample sizes specified in Paragraph 25 would be disproportionate or unduly burdensome under the circumstances, that Party shall promptly raise the issue with the Requesting Party. To the extent a dispute remains concerning the sample sizes to be used after good faith negotiations have occurred, either Party may request the Court's assistance in resolving such dispute.

27. The sample of documents comprised of the documents from Subsamples D(1), D(2), and D(3) shall be combined into a single Validation Sample, with no indication of the Subcollection from which the documents were derived or how they were previously coded. The Validation Sample shall be reviewed and coded by SME(s) who are knowledgeable about the subject matter of the litigation. This should be attorney(s) who is familiar with the document requests and the issues in the case. During the course of the review of the Validation Sample, the SME(s) shall not be provided with any information concerning the

Subcollection or Subsample from which any document was derived or the prior coding of any document. The intent of this requirement is to ensure that the review of the Validation Sample is blind; it does not preclude a Party from selecting as SMEs attorneys who may have had prior involvement in the original review process.

VI. Method of Recall Estimation

28. Pursuant to the method of recall estimation described herein, if the validation confirms the review achieved or exceeded the targeted recall rate, which in this case is a minimum of at least 70% recall, the TAR project is considered to be validated and review will cease. Note that the recall, as an absolute number in its own right, shall not be dispositive of whether or not a review is substantially complete. The Parties shall mind the novelty and materiality (or conversely, the duplicative or marginal nature) of any responsive documents identified in Subsamples D(2) and/or D(3). A recall estimate does not necessarily indicate that a review is inadequate, nor does a recall in this range or higher necessarily indicate that a review is adequate; the final determination also will depend on the quantity and nature of the documents that were missed by the review process.

- a. The number of responsive documents found = the size of Subcollection C(1) x the number of responsive docs in Subsample D(1) ÷ 400
- b. The number of responsive documents coded incorrectly = the size of Subcollection C(2) x the number of responsive documents in Subsample D(2) ÷ 400.
- c. The number of responsive documents not reviewed due to TAR = size of Subcollection C(3) x the number of responsive documents in Subsample D(3) ÷ 1,200.
- d. Estimated recall = the number of responsive documents found ÷ (the number of responsive documents found + the number of responsive documents coded incorrectly + the number of responsive documents not reviewed due to TAR).

29. Once the coding in Paragraph 27 has been completed, the Requesting Party will notify the Producing Party, and shall provide to the Requesting Party/ies:

- a. The results of the estimated recall calculation set forth in Paragraph 28 above, including the calculation of each of the elements in Paragraphs 28(a)-(c).
- b. The Bates numbers of the responsive documents from Subsamples D2 and D3; and
- c. The Bates numbers of the responsive documents identified in the Last Batches of review identified in Paragraphs 20-22 above, that were not previously produced or identified for production.

30. The Producing Party shall disclose whether the identified documents were derived from the Last Batches of review or the Validation Sample but is not required to disclose whether the documents from the Validation Sample came from subcollection D2 or D3.

31. Once the Requesting Party has received and has had an opportunity to review the items described in Paragraph 29, the Parties shall meet and confer to determine whether or not the Parties agree that the recall estimate, and the quantity and nature of the responsive documents identified through the sampling process, indicate that the review is substantially complete.

32. If the validation confirms the review achieved or exceeded a recall rate of 70% recall, the TAR project is presumptively considered to be valid and review may cease. However, if validation confirms that the review has not achieved 70% recall, additional review will be performed and the validation sampling will be repeated. If a Party reasonably believes that it cannot obtain this level of recall despite reasonable and proportional efforts, the Parties agree to meet and confer regarding additional measures and/or alternate steps.

33. If the Parties are unable to agree on whether the review is substantially complete, or whether the validation process must be repeated, the Parties shall promptly bring the issue to the Court's attention.

VII. Confidentiality and Ongoing Cooperation

34. Any proposed deviations in the agreed-upon TAR process set forth above will be promptly disclosed to the Receiving Party so that the Parties may confer about whether such deviation is acceptable.

35. The Parties will cooperate throughout the TAR process and confer in good faith to resolve any disputes concerning this protocol before making such dispute the subject of an application to the Court.

36. The Parties further acknowledge that although precision, efficiency and effectiveness of the search methodologies are important, perfection is neither required nor possible.

IT IS SO ORDERED.

/s/ Brian E. Murphy

Hon. Brian E. Murphy
United States District Court Judge

DATED: July 21, 2026