Tag Predictions

How DISCO AI is Bringing Deep Learning to Legal Technology

A DISCO White Paper

By Alan Lockett, PhD, Trevor Jefferies, JD, Neil Etheridge & Alicia Brewer

DISCO

Tag Predictions

A DISCO White Paper

By Alan Lockett, PhD,
Trevor Jefferies, JD,
Neil Etheridge
& Alicia Brewer

DISCO's artificial intelligence (DISCO AI) platform introduces a new approach to predictive solutions in the legal market. This white paper will outline how DISCO AI for tag predictions embodies groundbreaking legal technology due to its state-of-the-art infrastructure, unique approach to continuous learning, and tested precision and recall metrics of its predictive model.

DISCO's Distinct Approach: Deep Learning

Continuous learning is disrupting the way technology-assisted review (TAR) is completed, doing away with a need for traditional seed or training sets. As the court wrote in the *Rio Tinto PLC v. Vale S.A.,* continuous active learning is replacing traditional concerns about seeds sets, making them moot.[1] Furthermore, seed sets may be subject to challenge and judicial scrutiny and therefore suffer from defensibility issues. However, most of DISCO's competitors still require reviewers to follow a strict process that disrupts one's preferred workflow in order to apply predictive coding. Additionally, traditional machine learning systems depend on having a large seed set reviewed by an experienced, senior attorney with deep knowledge of the case in order to generate accurate predictions, thus creating a bottleneck to initiating the overall review process.

DISCO's approach is different. We believe that the legal team should drive the review; the machine should sit in the passenger seat. Continuous learning is always on and always learning; it is continuous. Rather than tell the lawyer how to run a review, the system watches in the background like a legal assistant, learning how to predict the lawyer's tagging behavior. When the system has observed enough human review activity, it begins to provide tagging suggestions. It does so *asynchronously,* that is, on its own schedule, without the lawyer having to do anything other than turn on a switch. As the lawyer corrects or accepts these suggestions, an understanding of the review increases, helping the lawyer to structure his/her workflow more efficiently. Continuous learning tag suggestions allow the lawyer to develop a review workflow without the need for seed sets, enabling a flexible review strategy for every case.

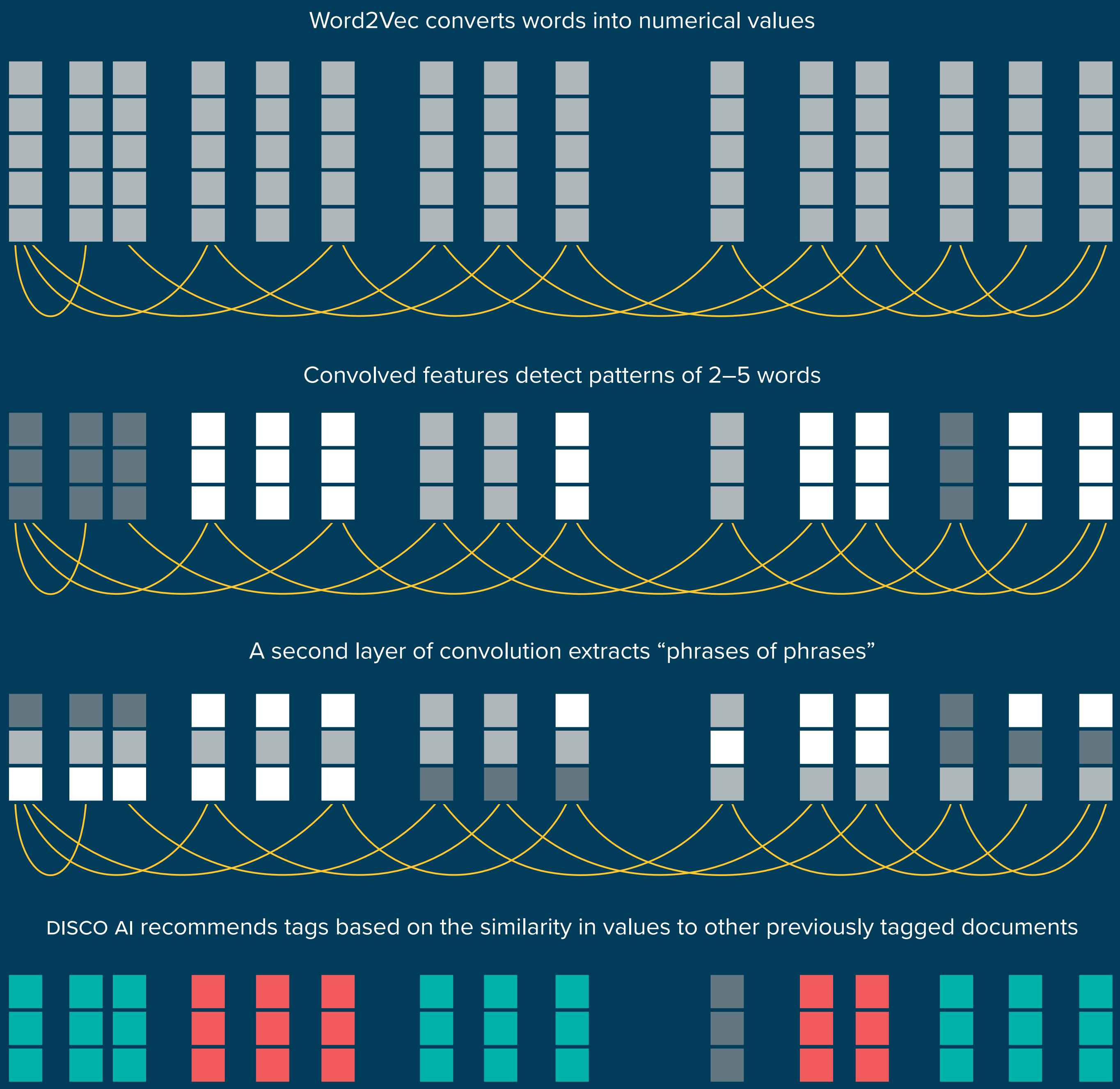
Underneath the hood, DISCO parlays words into meaning using a revolutionary tool called Word2Vec. Word2Vec was developed at Google to convert words to numbers in a way that encapsulates the immediate context around the word. Because words with similar meanings often occur in similar contexts, Word2Vec is able to extract the meaning of words to an astonishing degree.

DISCO's Convolutional Neural Network runs on top of Word2Vec in order to pinpoint key building blocks used to develop tag recommendations. Many competitors use a bag-of-words (BoW) model that simply counts how many times each word appears in a document, throwing away the word order. DISCO instead uses modern sequence processing techniques to read each word in the document in order to identify key phrases for predicting tag decisions.

DISCO's AI system, in contrast to most others, understands that the phrases *man bites dog* and *dog bites man* are very different, whereas the BoW model would find them identical.

Each convolutional layer in DISCO's CNN finds abstract phrases that match words by meaning. It does so by sliding multiple pattern matchers over the document; each pattern matcher discovers occurrences of key phrases. Higher layers arrange multiple phrases together in order to identify the crucial components of each tag being predicted. An operation called max pooling then shifts matched phrases around slightly, skipping extra words or words that would normally impede matching phrases. Once these complex phrase patterns have been located in the document, DISCO AI uses a traditional artificial neural network in order to suggest whether a tag should or should not be applied. All together, these resources power DISCO's AI technology to learn continuously and asynchronously. It doesn't interrupt the legal professional, but it's there when it is needed.

DISCO AI's 2-layer Convolutional Neural Network



The numbers produced by Word2Vec can be used in algebra-like statements to encode analogies.
If we take the numbers for *king* and subtract the numbers for *queen*, we get a set of numbers that can be added to *man* to produce the numbers for *woman*.

i.e., *king - queen = man - woman*

That is, Word2Vec understands differences of meaning, like the difference between male and female. Further interesting algebra-esque statements known to Word2Vec are

Russia + River = Volga

and

New York Yankees - New York = Boston Red Sox - Boston.

By converting words into Word2Vec numbers, DISCO AI understands the semantic context, e.g. a river in Russia or a sports team, of the documents word by word.

Results: Measuring DISCO AI

When looking at the real-world results of DISCO AI, we utilize four key metrics: positive accuracy, negative accuracy, recall, and enrichment.

Positive accuracy measures how often a tag recommended by DISCO AI is confirmed correct by a human applying the tag to a document.

Negative accuracy measures how often DISCO AI recommends a tag not be applied to a document and is confirmed correct by a human choosing not to apply the tag.

Recall compares the number of documents with a specific tag applied vs. the number of those documents for which DISCO AI suggested the tag.

Enrichment compares the prevalence of a particular tag being applied in the course of a traditional review over the prevalence of the tag being recommended by DISCO AI, if leveraging a DISCO AI workflow.

Positive and Negative Accuracy Results

DISCO AI shows impressive accuracy across many different types of tags. Looking at a sample of limited release DISCO cases, not only do responsive and issue tags perform well, great results are observed from confidential tag recommendations, such as privilege, and importance tags recommendations, such as hot. In the table below, see a breakdown of the average scores of positive and negative accuracy.

Tag Group	Positive Accuracy	Negative Accuracy	Recall
Confidential	77.63%	87.55%	82.44%
Issue	86.76%	96.78%	81.36%
Responsive	81.32%	71.86%	78.25%

Time to Accuracy Results

If one took a cross-section of the highest scores, the system is quickly surpasses human accuracy, often scoring in the high 80% and above. When reviewing this data, one of the key questions to ask is how many documents need to be tagged to reach this accuracy? Of course, the smaller the number, the more efficient the review can become. To take an example of each category above, the table below shows the tag type as well as the number of documents required to reach the listed level of accuracy.

Tag Type	Positive Accuracy	Tagged Documents
Privileged	75.52%	685
Responsive	89.16%	625
Issue	72.87%	137
Importance	76.70%	135

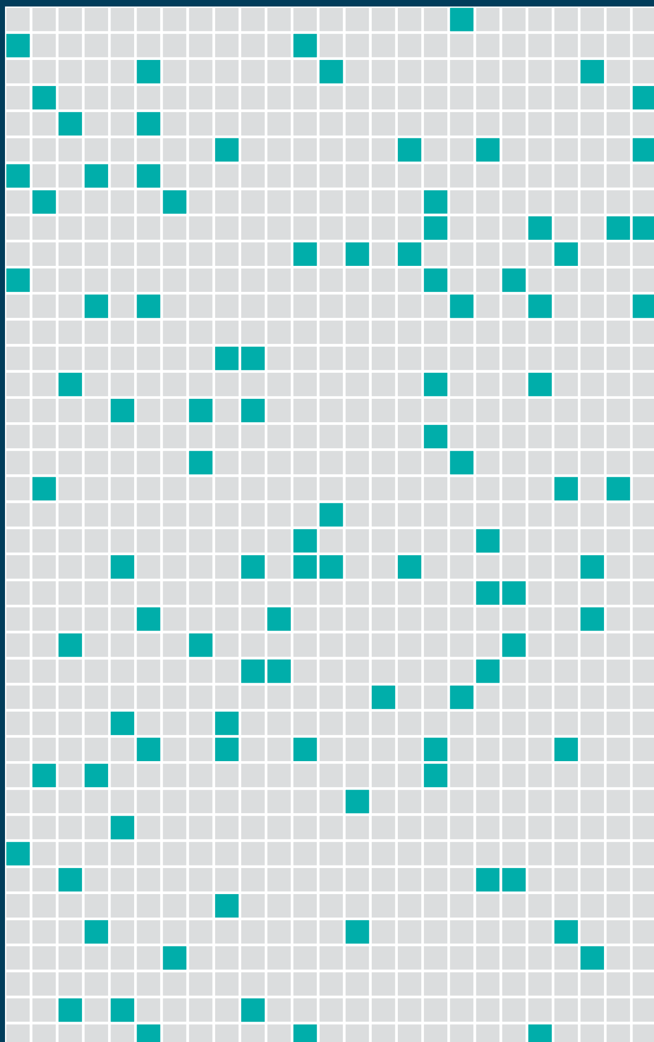
Improvement of Review with DISCO AI

As part of our limited release, we measured what improvements could have been seen had traditional document reviews leveraged DISCO AI. This measure, enrichment, compares the prevalence of a particular tag being applied in the course of a traditional review over the prevalence of the tag being recommended by DISCO AI, if leveraging a DISCO AI workflow.

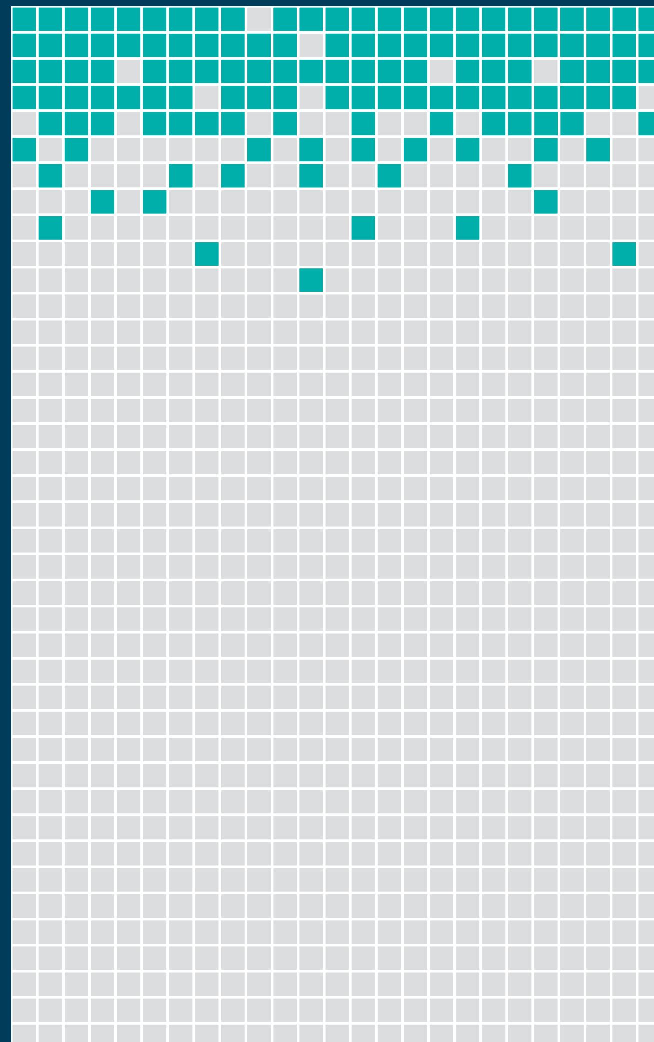
The table below illustrates the improvements achieved using DISCO AI in three test cases with different document counts in their respective review sets. From this information, we can see that using DISCO AI resulted in at least a 255% increase in efficiency, with much higher efficiencies reached in the larger matter. For example, in the largest matter, the review teams would have taken one-fifth the time for responsive review.

	Tag Type	Enrichment
4,238 Documents	Importance	2.61×
	Issue	2.55×
	Responsive	5.80×
392,045 Documents	Importance	11.15×
	Issue	78.62×
	Responsive	3.24×
895,918 Documents	Confidential	4.35×
	Issue	174.23×
	Responsive	4.96×

Natural Document Distribution



DISCO AI Prioritized Review



A DISCO AI prioritized review empowers reviewers to find the relevant documents much faster than with traditional, linear review where a natural distribution of relevant documents, as seen above, occurs. When running a structured review, DISCO AI can be combined with DISCO's Just-In-Time Batching to ensure that each new batch will have documents with the highest possible scores for the issue at hand amongst the remaining, unreviewed documents. In other words, DISCO AI enables you to automatically front-load batches with more relevant documents.

Case Study: DISCO AI Internal Test Scenario

To see the benefit of applying DISCO AI into a prioritized review, DISCO's AI algorithm was put through an internal test scenario that utilized data from the full, publicly available Enron data set, consisting of over 400,000 documents. DISCO's in-house legal team storyboarded a hypothetical lawsuit where Enron was sued by its shareholders, alleging mismanagement by its Board and Senior Executives. The principal allegations were:

- Enron dropped its focus away from its core domestic oil, gas and energy exploration and trading by focusing on overseas ventures.
- Enron's venture into broadband was ill-considered (and possibly due to self-dealing among the directors and/or senior staff), and again caused the company to lose focus on what should have been its core operations (i.e., domestic oil, gas, and energy).
- Enron was further mismanaged in that the company turned a blind eye or sometimes colluded in inappropriate conduct by employees and further did not monitor their considerable non-work-related use of Enron time and resources.

The DISCO review team generated coding decisions prioritized for documents that were not related to Enron work activities and documents that contained inappropriate, non-work related discussions or material in accordance with the review scenario. As a secondary experiment, DISCO also wanted to test the asynchronous, passive learning capability of DISCO AI to categorize documents, which was accomplished by coding for three work-related categories: “Enron Oil, Gas or Energy Business”, “Broadband”, and “Enron Outside USA.” A supervised review of 20,000 randomly chosen documents was conducted, in order to check the quality of both focused and passive DISCO AI tag predictions during the course of the review.

In regards to review prioritization, efficiency is key. For purposes of this test, efficiency was defined as the percentage of documents tagged in a given batch of documents, e.g. the accuracy on the top-ranked documents according to DISCO AI. A rate of 9.9% prevalence of non-work-related documents was found in the entire Enron dataset using prevalence sampling at the outset of review. DISCO AI was then able to achieve 55% efficiency for the tag “not work related” after reviewing only 100 positive signals. Using the prevalence of the tag in the tested corpus, we calculate this is a 5.6× enrichment over a traditional, non-AI review. Within 357 positive signals, the efficiency jumped to 81%. In combination with DISCO’s just-in-time batching, each subsequent review batch pulled for review would continue to increase the review’s efficiency. That is, as DISCO AI observed and continued to develop recommendations, its new learnings would inform the documents generated for each new batch as they were requested by a reviewer.

Tags not explicitly being trained were also able to provide impressive results. Concerning the tag “Enron - Oil, Gas or Energy” which had a 10.5% prevalence within the Enron dataset at large, it took 41 positive signals to obtain 42% efficiency. More so, 132 positive signals increased the efficiency to 80%. The coding completed against the primary tag “not work-related” was, by proxy, driving other concurrent tag learning. This passive learning was purely a contingent gain, made possible through DISCO AI concurrent tag learning capabilities.

Concluding Statements

Within the context of legal proceedings, several landmark rulings, not the least of which was *Rio Tinto PLC v. Vale S.A.,* confirm the rise in use of AI technology in ediscovery.[2] However, “its widespread application — and the realization of its potential benefits — has been impeded by uncertainty: about its acceptance by the courts as a legitimate alternative to costly, time-consuming manual review of documents in discovery.”[3] Nevertheless, several cases “reflect the parties’ use of TAR, without otherwise addressing its use.”[4] Thus, best practices must be considered when implementing a review strategy, regardless of the technology used or eschewed. While many courts and commentators agree that technology assisted review (TAR) should be held to the same standard of reasonableness as any other discovery process,[5] because no review is the same, and case or jurisdictional requirements will vary, attorneys will need to determine the reasonableness for using (or not using) AI for any particular case. Should one decide to use AI for document review, per the data stated above, we believe DISCO provides a leading solution that is defensible, efficient, accurate, and will save time and money on every case.

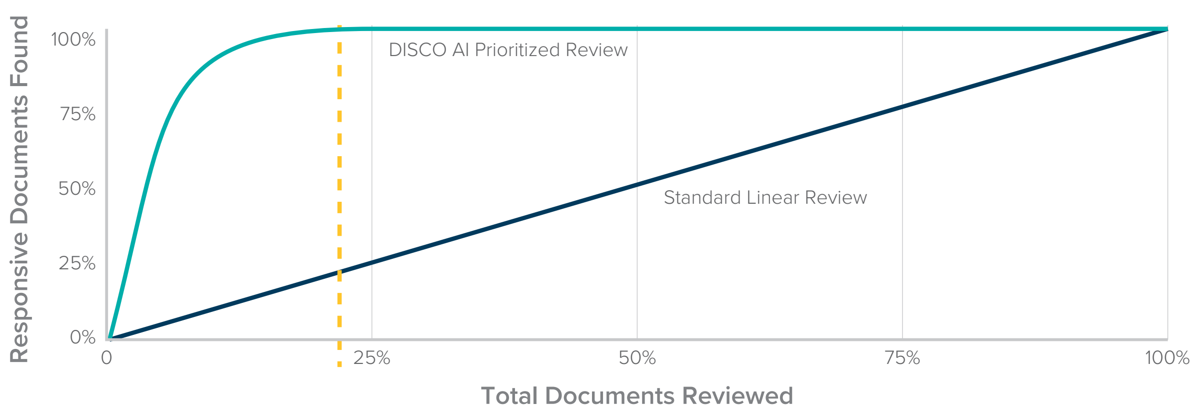
Appendix: DISCO Best Practices and Recommendations

This appendix outlines a simple review process using DISCO AI that may provide some insight for any particular case. The process includes randomly sampling the set of documents to be culled, culling and/or mass tagging to winnow down the potential set of responsive documents, randomly sampling that set for a prevalence estimate of particular tags and quality control (QC), performing the review using a combination of DISCO AI and more traditional keyword searching, followed by a final sampling to ensure the results are acceptable.

The following hypothetical case will provide more detail to this process: Assume a set of data has been collected from the client as potentially responsive to requests for production from an opposing party. After de-duping, de-NISTing, etc., the remainder of the data yields a corpus of 1.1 million documents. A cursory "macro" review (e.g., using document types, date ranges, or common email spam domains) yields 100,000 documents as clearly non-responsive and these are removed from the corpus.[6] At this point, there are 1 million documents remaining that are potentially responsive and need to be evaluated in more detail.

The next step is to randomly sample the documents to get a baseline of the number of documents that are responsive (that is, the "prevalence"). To achieve a 95% degree of confidence with a 2% margin of error,[7] a random sample of 2,395 of the remaining 1 million documents would need to be reviewed (that number can be found using any one of many online sample calculators, or using DISCO's software). After reviewing the 2,395 random sample set of documents, the review manager would then have their target range of likely responsive documents in the 1 million document population. For example, assuming one found 17% of the sampled documents as responsive, that would mean that one could anticipate that between 15–19% (or between 150,000 and 190,000) of the underlying population would be responsive. In fact, one can say that they are 95% certain of their range, which was the "confidence" level provided by the sample.

With those numbers in mind, one can begin the review, using DISCO AI along with any one or more of the traditional methods. One suggestion is to begin by doing "obvious" or "precise" keyword searches or search strings, such as the fairly unique name of the project, product, or contract that is at issue in the litigation, or a linear review of the most critical dates or custodians, and sorting those search results using DISCO AI. After the lawyer has exhausted these obvious methods, begin reviewing according to the DISCO AI predictions of responsive documents. DISCO AI provides a score for each document, so one could sort the entire database and review those documents that DISCO's AI rates as the "most likely" to be responsive — in ranked order according to the score. Using a managed review, DISCO's AI can be combined with DISCO's Just-In-Time Batching to ensure that each new batch that is checked out by a review team member will have documents with the highest possible AI scores for the remaining unreviewed documents.



When the number of reviewed documents reaches the target prevalence range for responsiveness (for example, 155,000 responsive documents have been found), and after the algorithm no longer recommends any additional documents, (e.g. the predictive ranking shows that no more responsive documents exist) consider taking a second random sample, this time of the remaining unreviewed documents. Again, let's assume for round numbers that to find the 155,000 responsive documents, one also found 115,000 non-responsive documents in the course of the review; thus leaving 730,000 documents that have not been reviewed at all.

For the random sample of the unreviewed set, the review manager would probably want a higher degree of confidence and lower margin of error than their initial sample, since they may need to use this second sample to defend their work. An acceptable number might be a 99% confidence level, with a 2% margin of error, which would require in this case a random sample of 4,137 of the 730,000 "population" of the unreviewed documents. Let's assume the lawyer found that approximately 1% of the sample was in fact responsive (that is, 41 documents in the sample were responsive).

With those numbers in mind, the question is what to do? Should the review continue? Can one defend a decision to stop reviewing? Of course, the answer is it depends, and it cannot be overemphasized that this decision should be based on the legal judgment of the lawyer managing the review. The most basic analysis would be that there are (with 99% confidence) no more than 10,658 of the 730,000 unreviewed set that are responsive. Using the metrics ascertained in the review to provide the approximate number of documents that can be reviewed per hour (that is, to review the set to get to 155,000 responsive), the approximate cost of reviewing the additional documents is fairly easy to quantify. For example, assume that a review group reviewed 50 documents per hour, with an average hourly rate of $50 per hour. To review the remaining 730,000 documents would then cost approximately $730,000. Much harder to quantify, of course, is the potential “benefit” that (in all likelihood the opposition would argue) the remaining review might yield. If the entire amount in controversy is $100,000, the proportionality analysis is “probably” straightforward (and in fact the entire scope of this review would have been questionable).

However, a proportionality analysis may not be appropriate until all avenues of review have been exhausted except for a full linear review. That is, if keyword, date, custodian, or other metadata searches could reasonably target some or all of the remaining 10,658 responsive documents, those efforts should also be evaluated. One simple method is to use the 41 documents found in the second sample, and determine if these 41 documents suggest any other avenues by which more responsive documents could be identified. Similarly, but with more effort, information learned during the review of the 155,000 responsive documents may provide additional clues for searching the remaining corpus of unreviewed documents. A defensibility position needs to anticipate the argument that there is a “better” (and cheaper) alternative to a full linear review; namely that a targeted search would significantly reduce the cost component of a given proportionally analysis. Once those potential objections are addressed, counsel will at least have the ammunition necessary to defend the decision to stop the review.

And speaking of defensibility, it is important to document the decisions made during the workflow. Maintain records and lists of any keywords, custodians, date ranges, etc. used for culling decisions, what sample calculations and calculators were used and results, prevalence estimates found for each measured issue (e.g. privilege, responsiveness, issues), and alternative search strategies and results of each. With the combination of powerful technology such as DISCO AI and documented statistically-accepted methodologies, counsel will be able to maximize search potential while providing the client with the most cost-effective review.

Notes

[1] *Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125, 128 (S.D.N.Y. 2015) (citing Gordon V. Cormack & Maura R. Grossman, *Evaluation of Machine Learning Protocols for Technology-Assisted Review in Electronic Discovery,* in Proceedings of the 37th Int'l ACM SIGIR Conf. on Research & Dev. in Info. Retrieval (SIGIR '14), at 153–62 (ACM New York, N.Y. 2014), http://dx.doi.org/10.1145/2600428.2609601; Maura R. Grossman & Gordon V. Cormack, Comments On "The Implications of Rule 26(g) on the Use of Technology–Assisted Review," 7 FED. CTS. L. REV. 285, 298 (2014) ("Disclosure of the seed or training set offers false comfort to the requesting party . . . .")).

[2] *See, e.g., Rio Tinto,* 306 F.R.D. at 129.

[3] *The Sedona Conference TAR Case Law Primer,* at iii (Public Comment Version, August 2016). The Sedona Conference Working Group Series (WG1) (available at https://thesedonaconference.org/download-pub/4812)

[4] *Id* at 8.

[5] See, e.g., Rio Tinto, 306 F.R.D. at 129; The Sedona Conference, Commentary on Defense of Process, at 32-33 (public comment version, September 2016) (available at https://thesedonaconference.org/publication/sedona-conference-commentary-defense-process-public-comment-version-september-2016).

[6] One might also choose to remove from the predictive workflow any documents that may not lend themselves well to the applicable predictive technology, such as file types with predominantly graphic images or numerical data, or even foreign language if the predictive technology does not accommodate foreign language.

[7] Parties may choose to agree on a particular degree of confidence and margin of error, as did the parties in *Rio Tinto PLC v. Vale,* 306 F.R.D. 125 (S.D.N.Y. 2015) (agreeing on a 95% degree of confidence and a 2% margin of error).

DISCO