Tag Predictions

How DISCO AI is Bringing Deep Learning to Legal Technology

By Alan Lockett, Ph.D., Liad Levinson J.D., Robert Harrington, Ph.D., Anush Emelianova, J.D.

DISCO

Tag Predictions

A DISCO White Paper

DISCO's artificial intelligence (DISCO AI) introduces a new approach to predictive solutions in the legal market. This white paper outlines 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.,* the use of continuous learning eliminates traditional concerns about seed sets (the initial set of documents used to train an AI model).[1] As data is processed, the model continues to learn and improve. Seed sets may be subject to challenge and judicial scrutiny and therefore suffer from defensibility issues. However, many of DISCO's competitors still require reviewers to follow a strict process that disrupts one's preferred workflow in order to apply predictive coding. 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 and the machine should sit in the passenger seat. Continuous learning is always on, allowing it to learn and improve constantly without interrupting the review process. 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 structure their workflows 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 fastText. fastText was developed by Meta's AI research lab 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, fastText is able to extract the meaning of words to an astonishing degree.

DISCO's convolutional neural network (CNN) runs on top of fastText 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.

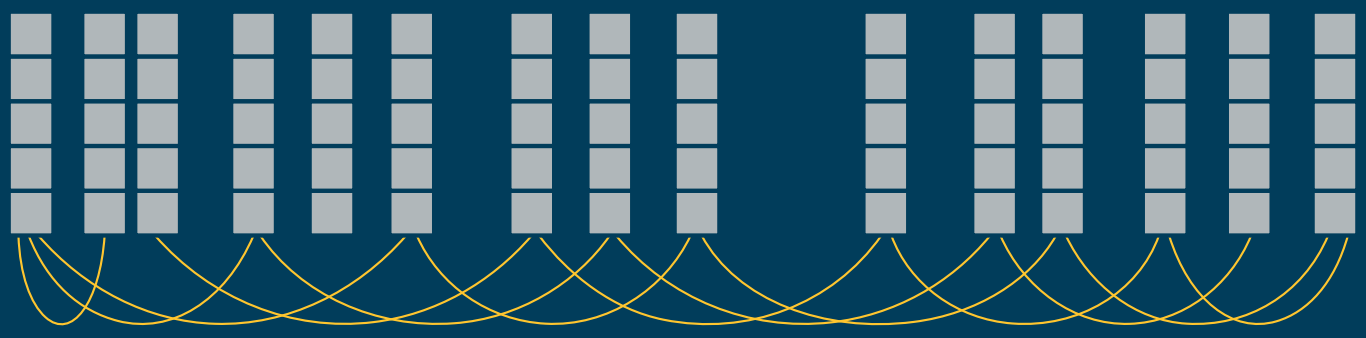
DISCO's AI system, in contrast to most others, understands that the phrases *man bites dog* and *dog bites man* are very different, whereas the BoW model would find them identical.

DISCO's CNN can be analogized to CNNs used in image recognition technology, such as facial recognition. When a CNN examines an image to determine if it contains a face, it moves through different layers at increasing levels of complexity. For example, at one of its lowest levels, a CNN will look for certain textures or colors that often depict a person. At a higher layer, the CNN looks for details like the shape of a nose or ears. Higher layers will build on previous layers to determine if a face is in a frame.

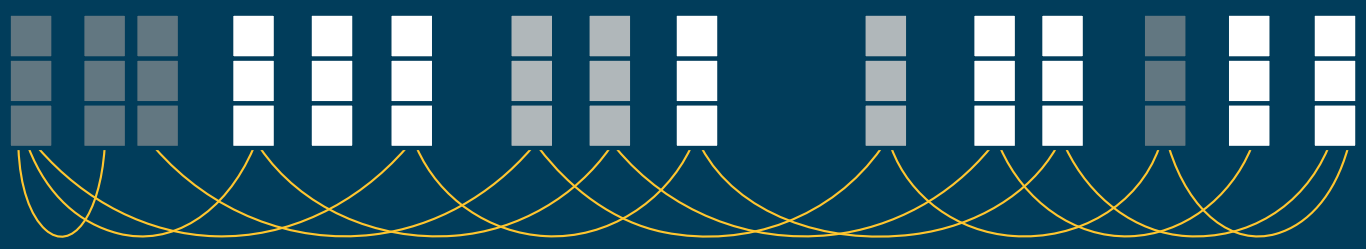
DISCO's CNN follows a similar process consisting of two layers. At the first layer, our CNN tries to find a few words that could be responsive to a specific tag. As we move to the second layer, our CNN grows its focus from individual words to sentences or phrases that could be responsive. By breaking down documents layer by layer, we can consider the order and relationship words have to one another, allowing our AI model to accomplish much more than something like a key phrase searcher.

DISCO AI's Two-layer Convolutional Neural Network

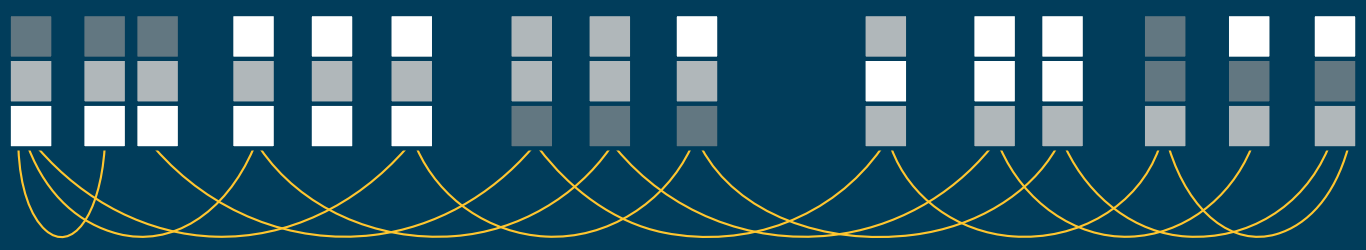
fastText converts words into numerical values



Convolved features detect patterns of 2–5 words



A second layer of convolution extracts "phrases of phrases"



DISCO AI recommends tags based on the similarity in values to other previously tagged documents

Reading Words as Numbers with fastText

The numbers produced by fastText can be used in algebra-like statements to encode analogies. If we take the numbers for *Paris* and subtract the numbers for *France*, we get a set of numbers that can be added to *Poland* to produce the numbers for *Warsaw*.

i.e., *Paris - France = Poland - Warsaw*

That is, fastText understands differences of meaning, like the difference between Paris and Warsaw. Further interesting algebra-esque statements known to fastText are

Russia + River = Volga

and

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

By converting words into fastText numbers, DISCO AI understands the semantic context (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 several key metrics: accuracy, recall, and enrichment.

Accuracy measures how often a tag recommended by DISCO AI is confirmed correct by a human applying the tag to a document. Accuracy can be positive (correctly applying a tag — also known as precision) or negative (not applying a tag).

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

Enrichment measures how much more frequently documents that should receive the relevant tag appear in DISCO AI-prioritized documents compared to the prevalence of such documents in the general document population.

Accurately Identifying Responsive Documents with DISCO AI

Using DISCO Ediscovery with built-in AI and best-practice workflows, attorneys should be able to reduce document populations with high positive accuracy and efficiency.

In one review involving 11,000 documents, DISCO AI demonstrated significant accuracy when compared against manual review — which can typically achieve approximately 70% positive accuracy and 70% recall.[2] The documents had previously been coded by associates at an Am Law 50 firm, but DISCO performed the same review as a proof of concept. DISCO AI correctly identified over 99% of the documents that it deemed "highly likely" to be responsive or non-responsive — when compared to the work of the Am Law 50 associates. For documents that DISCO AI deemed "likely" to be responsive or non-responsive, Am Law reviewers agreed approximately 90% of the time — validating DISCO AI's confidence level.

Tag Prediction	Confidence	Accuracy	Tagged Documents
Responsive	Highly Likely	99.00%	3,627
	Likely	89.70%	2,511
Non-Responsive	Highly Likely	99.20%	2,210
	Likely	93.30%	1,492

Achieving Recall with DISCO AI

While DISCO AI makes tagging suggestions, it does not itself apply tags. In the same review discussed above, DISCO contract attorneys were shown DISCO AI's tag suggestions but ultimately made their own calls on responsiveness when applying tags. Across the document population, the tags applied by the DISCO reviewers were almost identical to those applied by the associates at the Am Law 50 firm. Where there were differences, the client determined that the documents represented edge cases and it was reasonable for the calls to have gone either way. This example demonstrates near-perfect recall because DISCO AI was able to quickly and accurately find all prima facie responsive documents in the population.

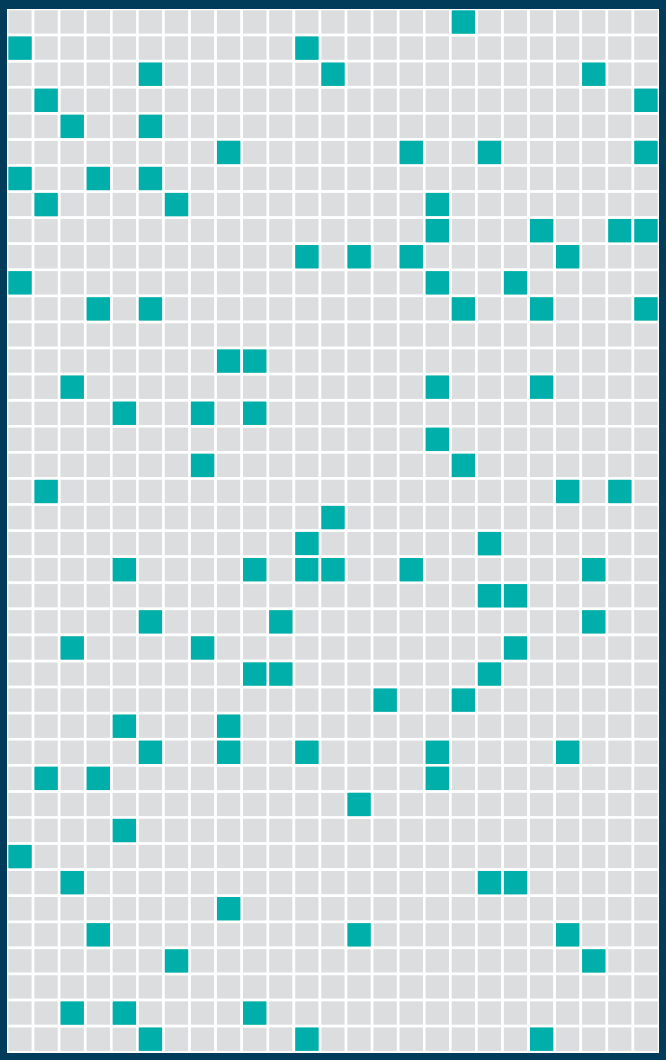
Enrichment of Review with DISCO AI

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 when leveraging a DISCO AI workflow.

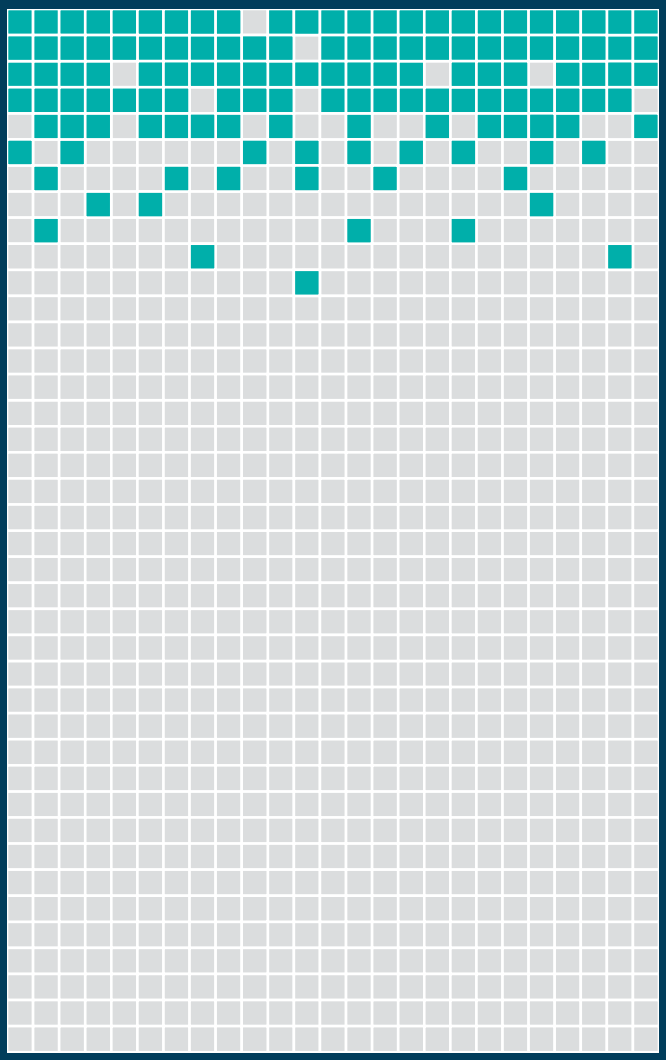
The table below illustrates the improvements achieved using DISCO AI in three test cases with different document counts in their respective review sets. From this information, we can see that using DISCO AI resulted in at least a 255% increase in efficiency, with much higher efficiencies reached in the 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 a traditional, linear review with a natural distribution of relevant documents. 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." DISCO's review team conducted a supervised review of 20,000 randomly chosen documents to check the quality of both focused and passive DISCO AI tag predictions during 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. The tag "Enron Oil, Gas, or Energy," which had a 10.5% prevalence within the Enron dataset at large, took 41 positive signals to obtain 42% efficiency. 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, many 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.[3] 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."[4] Nevertheless, several cases "reflect the parties' use of TAR, without otherwise addressing its use."[5] 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,[6] 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.

For those choosing to use AI for document review, we believe the provided data indicates that DISCO offers a leading solution that is defensible, efficient, accurate, and will save time and money on most cases.

Appendix

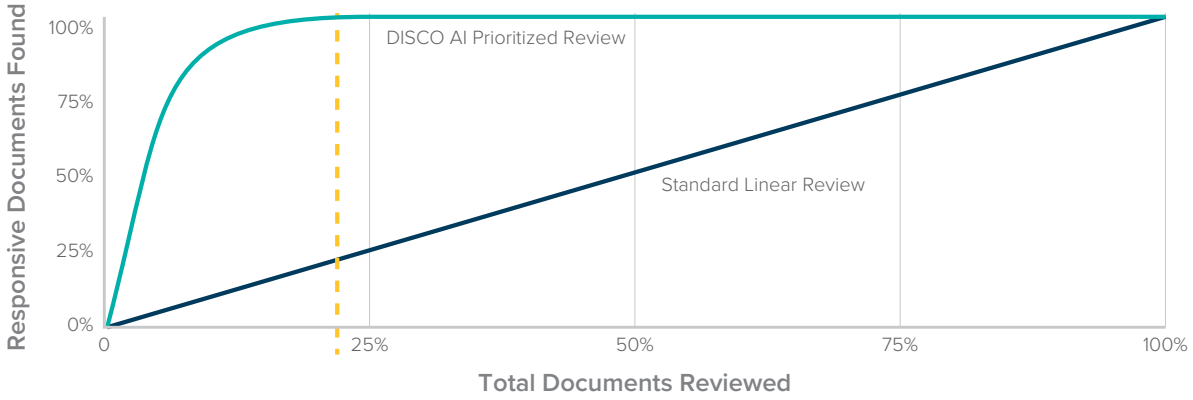
DISCO Best Practices and Recommendations

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.[7] 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,[8] 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 documents of the set of 730,000 unreviewed documents 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 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.

For 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] Gordon V. Cormack & Maura R. Grossman, *Navigating imprecision in relevance assessments on the road to total recall,* at 9, Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information (August 2017), available at https://dl.acm.org/doi/10.1145/3077136.3080812 (noting that human assessor can achieve on the order of 70% recall and 70% precision).

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

[4] *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/down-load-pub/4812).

[5] *Id* at 8.

[6] *See, e.g., Rio Tinto,* 306 F.R.D. at 129; *Winfield v. City of New York,* No. 15-CV-05236LTSKHP, 2017 WL 5664852, at *9 (S.D.N.Y. 2017) (holding that "in the absence of evidence of good cause," there is no basis for "courts to insert themselves as super-managers of the parties' internal review processes, including training of TAR software, or to permit discovery about such process"); 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.

[7] 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.

[8] 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).

About the authors

Dr. Alan Lockett is the VP of Machine Learning and AI at DISCO. He received his PhD in Computer Science from the University of Texas at Austin.

Liad Levinson is the Senior Director of Product Domain at DISCO. He received his JD from Columbia Law School and previously practiced as a litigation attorney.

Dr. Robert Harrington is the Director of Machine Learning and AI at DISCO. He received his PhD in Physics from Northeastern University.

Anush Emelianova is a Product Marketing Manager at DISCO. She received her JD from Harvard Law school, and previously practiced as a privacy and litigation attorney.

DISCO (NYSE:LAW) provides a cloud-native, artificial intelligence-powered legal solution that simplifies ediscovery, legal document review and case management for enterprises, law firms, legal services providers and governments. Our scalable, integrated solution enables legal departments to easily collect, process and review enterprise data that is relevant or potentially relevant to legal matters.

www.csdisco.com