

The Full Potential of Technology Assisted Review

In their 2012 RAND report, “Where the Money Goes”, Nicholas Pace and Laura Zakaras point to predictive coding as an essential solution to the burden of rising document review costs in ediscovery. Predictive coding, also referred to as technology assisted review (“TAR”), is a combination of technology, people, and process that promises to replace much of the costly linear review occurring today. But TAR can do more than just replace traditional linear review, and learning how to use TAR to its full potential will lead to cost savings and efficiency gains in unexpected areas. This paper provides an overview of TAR and explores some alternative uses of this very powerful tool.

TAR Explained

The most widely understood definition of TAR is the application of supervised machine learning techniques to ediscovery review workflows. At the heart of this process is a subject-matter expert training a computer system to distinguish between relevant and non-relevant documents. While the computer system’s underlying statistical algorithms are essential to this endeavor, the technology itself is useless without a human expert following a precise training process that includes random sampling and rigorous validation efforts. A good way to illustrate the fundamentals of TAR is to compare it to traditional linear review.

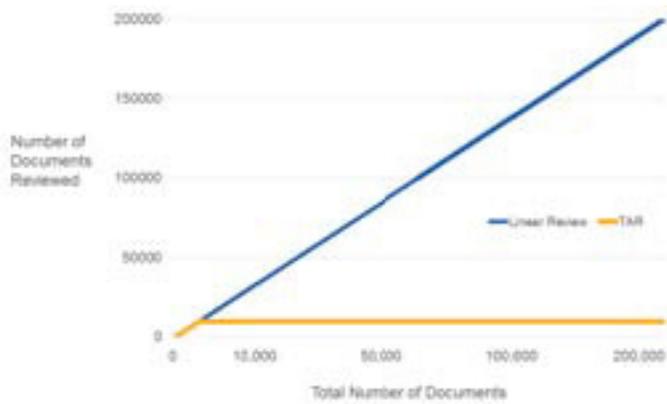
A typical linear review begins with the application of search terms to reduce the document set to manageable levels. The resulting search hits are then divided into batches for review by a team of review lawyers. The

reviewers apply coding decisions to each document in their batch, and anything coded as relevant is either moved on to second-pass review or produced directly to the requesting party.

When using TAR to replace linear review, a subject matter expert (or small team of experts) is selected to train the computer system. The entire document set (not reduced by search terms) can be fed into the system, and a series of random document samples is presented to the expert for review. It is also possible to use TAR after search terms are applied. The expert’s decisions are used by the system to create a model of relevance, and each training iteration refines the model further. Once the model has achieved a high level of accuracy, it is used to make a relevance decision on every document in the set. Random samples from both the relevant and non-relevant sets are then reviewed to validate the system’s coding. Once validated, any document coded as relevant is either moved on to second-pass review or produced directly to the requesting party, as it would in a linear review. Typically, the expert needs only review a few thousand documents to achieve optimal results, regardless of the size of the data set.

In effect, TAR amplifies the expert’s review efforts. By reviewing only a few thousand documents, the expert gets consistent, validated coding decisions across the entire document set. It has been demonstrated through numerous studies that TAR results match or exceed the results of a linear review and are achieved in a fraction of the time and at a fraction of the cost. The chart below illustrates the potential cost savings provided by TAR.





Corporations and law firms who have used TAR have reported varying degrees of success. In each case where we encountered a reported lack of success, it was also clear that there was a lack of expertise. For TAR to be successful, it is essential that it be undertaken by a team with expertise. The subject matter expert, of course, must have an extensive understanding of the data set and the relevant issues. Beyond that, a successful TAR project must be managed by a team of people with a deep understanding of technology, statistics, and ediscovery workflows. It's important to have someone who understands the underlying machine learning algorithms so that potential training problems can be identified early and fixed before results get skewed. The team should also include someone comfortable with the nuances of statistical sampling and the various means of measuring information retrieval success (e.g., precision, recall, and F-score). Finally, because every case has unique workflow requirements, the team needs real-world ediscovery experience to fully understand how TAR fits into the overall case strategy. Assembling such a team dramatically increases the odds of success. Furthermore, such a team would be adept at exploiting TAR's inherent flexibility.

TAR's Flexibility for Alternative Uses

The RAND report trumpets TAR as a solution to rising review costs because of its potential as a substitute for linear review. In fact, it is a tremendous tool that has proven adept at fulfilling this promise, dramatically reducing review costs in large data sets; however, it

would be a mistake to limit the use of TAR to just that task. TAR is a powerful tool that, with a bit of creativity, can be used in a variety of ediscovery situations. The remainder of this paper will highlight three alternative-use examples that demonstrate the power and flexibility of TAR.

1. INCOMING PRODUCTIONS

The exploding size of data doesn't just impact those parties producing documents, but also the parties receiving productions. When a party receives an incoming production, they know that all of the documents are ostensibly relevant, but they don't know how each document relates to the specific issues of the case. Traditionally, gaining this understanding required an issue review of the entire incoming production. As productions grow larger and larger, this type of review becomes cost and time-prohibitive. TAR can solve this problem.

Using TAR to issue code an incoming production requires extra care at the beginning of the process when selecting the issues around which to review. It's critical to choose issues that have minimal overlap with each other, or else the TAR system might have difficulty distinguishing one issue from another. For example, if the litigation revolves around marketing and finance functions at both the LA and NY offices, it makes more sense to use "marketing" and "finance" issues rather than "LA" and "NY" issues because the system can do a better job distinguishing marketing documents from finance documents than it can distinguishing marketing and finance documents created in LA versus those created in NY. Some TAR platforms, like Lighthouse's Clarity, are able to alleviate this problem by extracting more nuanced information from documents, but even then it's better to start with clearly defined issues that are as distinctive as possible.

Once issues are selected, an expert reviewer can start training the system by reviewing a random sample of documents, coding documents as they relate to one or more issues. Once training is complete, the system automatically identifies the documents associated with each issue, allowing the trial team to focus on clearly defined subsets rather than dedicating the time and expense to wade through the whole production.



2. IMPROVING LINEAR REVIEW

Linear review can be a costly and time-consuming process, but sometimes eyes-on review is necessary. A good example of this is potentially-privileged documents—most parties will insist on an eyes-on review of potentially privileged documents. In these situations, TAR can be used to enhance the efficiency and accuracy of linear review. Using TAR in this way requires a TAR platform, such as Lighthouse’s Clarity, that is designed to be able to run in the background of a linear review. During the course of the review, as humans make coding decisions on their review batches, these coding decisions are fed into the TAR system. The system can then be used to prioritize review of relevant documents and act as an automatic QC mechanism.

Document review is part of a workflow whose purpose is the production and presentation of relevant documents. This workflow works best when relevant documents are identified as early as possible. Earlier identification means that documents are ready to produce more quickly, decreasing the risk of missed deadlines. Earlier identification also means that key documents can be pushed to trial counsel more quickly, increasing their usefulness in shaping strategy.

In a traditional linear review, relevant documents are just as likely to be found in the last batch reviewed as the first batch, so document production won’t be substantially complete until review is substantially complete. Furthermore, early identification of key documents only happens by serendipity. In contrast, if a TAR system is trained in the background of a linear review, it can identify documents most likely to be relevant and push them forward in the review queue. This dynamic prioritization increases the odds of locating key documents more quickly, and allows the bulk of production to occur earlier in the review.

One of the common criticisms of linear review is the inevitable coding inconsistencies that arise when multiple reviewers are working on the same project.

Coding inconsistencies cast doubt on the integrity of a document production, and this doubt can be exploited for tactical advantage by the opposing party.

The problem is even worse when reviewers make inconsistent privilege decisions or protective order designations. These errors can result in potentially catastrophic outcomes such as the loss of attorney-client privilege or the public release of trade secrets.

A TAR system trained in the background of a linear review can alleviate this problem, increasing the accuracy and defensibility of linear review by acting as an automatic QC mechanism. The TAR system can detect inconsistent coding and outlier decisions, alerting the review manager so that the problems can be addressed. The result is a cleaner, more consistent document production that reduces the risk of costly errors.

3. NEEDLE IN A HAYSTACK

Companies are sometimes required to look for types of documents that might not even exist. If relevant documents don’t exist, or are few and far between, then a TAR system can’t be trained on relevance. But with a subtle change of perspective, TAR becomes extremely useful in this situation: instead of training on relevance, the system can be trained on non-relevance. In other words, the system is trained to cull out everything that’s clearly not important, leaving behind the set of documents that might be important.

For this approach to be effective, the expert reviewer should code each document in the training sample as either “not relevant” or “possibly relevant”. Before coding a document as “not relevant”, the expert must determine that any document with similar content couldn’t be relevant. If there is doubt about the relevance of similar documents, then the document should be coded as “possibly relevant”. Once training is complete, all documents coded by the system as “not relevant” can be removed from further consideration, providing an efficient and defensible culling technique.



Conclusion

TAR has proven useful for replacing eyes-on review, and is becoming widely adopted for that purpose. TAR's flexibility has created other uses that may become just as impactful over time. Whether it's sorting through incoming productions, enhancing linear review, or culling out non-relevant documents, TAR's usefulness as an ediscovery tool extends beyond just the replacement of linear review. TAR adoption is increasing at a rapid pace and, in one form or another, will soon become an integral part of every ediscovery workflow.

About Lighthouse

For 25 years, Lighthouse has provided innovative software and services to manage the increasingly complex landscape of enterprise data for compliance and legal teams. Lighthouse leads by developing proprietary technology that integrates with industry-leading third-party software, automating workflows, and creating an easy-to-use, end-to-end platform. Lighthouse also delivers unique proprietary applications and advisory services that are highly valuable for large, complex matters, and a new SaaS platform designed for in-house teams. Whether reacting to incidents like litigation or governmental investigations, or designing programs to proactively minimize the potential for future incidents, Lighthouse partners with multinational industry leaders, top global law firms, and the world's leading software provider as a channel partner.

Connect with us to see how Lighthouse can best support you.

(206) 223-9690 | lighthouseglobal.com | info@lighthouseglobal.com



LIGHTHOUSE