The fundamentals of Technology Assisted Review (TAR).

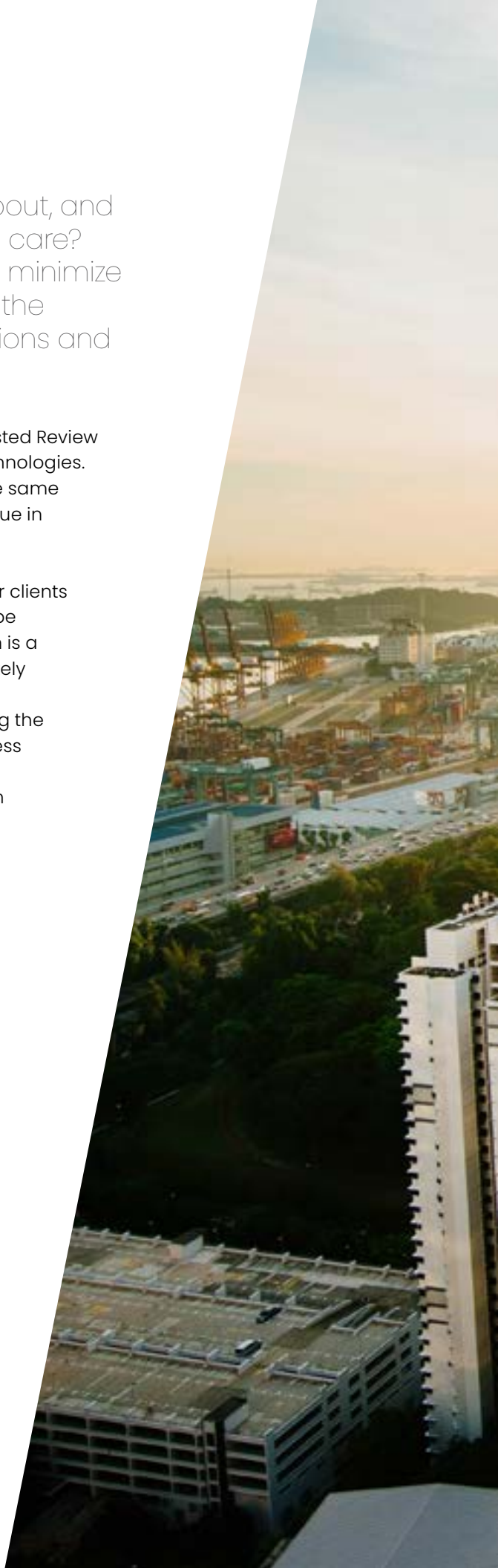
What is this "TAR" that everyone is talking about, and perhaps more importantly, why should you care? Does it really help speed up discovery and minimize costs? Is it a defensible process? What are the courts' views on TAR use? All of these questions and more will be answered within this paper.

First, it is important to understand what TAR is. Technology Assisted Review (TAR) is an umbrella category that covers many advanced technologies. There are many different TAR approaches but they all share the same goal: to help legal teams separate datasets by relevancy or issue in order to prioritize documents for expedited review.

TAR is generating a great deal of interest among law firms, their clients and corporations. Law firms, always cautious, can sometimes be reluctant to use TAR workflows because they fear the approach is a "black box" technology that they would not be able to adequately explain to a court if issues were to arise. They also fear that the expense will outweigh the return. Interestingly, clients are driving the use of TAR because it provides a principled framework to address massive volumes of data. Clients understand that TAR reduces the overall number of documents for review, which can result in significant cost savings.

The next step is to drill down into the two primary technology categories within TAR; namely, concept-based and support vector-based applications, which operate on different classification algorithms. Understanding the basics, even at a high level, is instrumental in choosing the right solution for a particular use case. TAR workflows can be levied for early data analysis, data reduction and culling and used to focus on specific, substantive datasets for discovery, investigations, depositions or other matter-specific objectives. Today, many terms are associated with TAR in the marketplace: keyword searching, clustering, categorization, themes, analytics, computer assisted review and predictive coding, among others. A full-blown discussion of all of these terms would be encyclopedic, but it is worthwhile to have a high-level understanding of each one.

This paper concludes by summarizing recent court decisions referencing TAR, determining how parties can use it defensibly and evaluating how it can help to achieve reasonableness and proportionality.



The history of TAR.

The identification, collection, review and production of Electronically Stored Information (ESI) has brought about a revolution in discovery, mainly because of the sheer volume of information that must be addressed. The first time a legal team used software to view images of scanned documents, technology was assisting the team's review. Even though technology was used, this was still considered a linear, document-by-document review with limited searching capabilities.

In the beginning, there was keyword searching.

Keyword searching – in which a pre-determined set of words are run against the metadata, extracted text and/or OCR of datasets and only documents that "hit" on one or more keywords are reviewed – was considered a breakthrough technique a few decades ago. Keyword searching further advanced in the 1990's through the use of Boolean searches, which allow combinations of terms to disambiguate over-inclusive keywords. For example, a search string would state "include (keyword1, keyword2, keyword3, etc.) but not (excludeword1, excludeword2, excludeword3, etc.)". These combinations of related words and concepts, or "ontologies," often include proximity limiters (i.e. Abraham w/2 Lincoln). Boolean searching improved the retrieval of relevant documents, but given the syntax rules required to formulate Boolean searches, litigation counsel had to engage highly skilled linguists to develop productive combinations. This was expensive and time-consuming, although it did create a short-lived burst of notoriety and exposure for linguistic experts. Some would say keyword searching was the beginning of Technology Assisted Review. While a number of tagging protocols and sorting features has developed, keyword searching remains the most commonly-used method for reducing the number of documents to be reviewed.

Then came categorization.

More recently, sophisticated categorization algorithms packaged within eDiscovery software exploded onto the scene as an efficient and cost-effective method of slicing and dicing enormous quantities of ESI. Categorization systems identify and label documents that pertain to a certain topic or issue, based on a set of exemplar documents.

Many different algorithms categorize ESI. Some applications use a conceptual indexing algorithm to categorize by concept, subject or legal issue, based on document exemplars fed to the system. Others such as email threads, near-duplicate detection, language identification and themes use syntactical algorithms to categorize documents.

Different classification algorithms power these systems. The models can be based on word indexes, dictionaries, thesauruses, the context within which words commonly occur throughout datasets and ontologies.

Many categorization tools operate with no human intervention or training and simply apply the algorithm to organize datasets into similar clusters or themes. For such applications, however, the clustering that results may not necessarily be helpful to the review team, because it is organized purely by the algorithm's classification. A more productive approach would start with human review of a set of documents in order to code or differentiate by relevancy, issue, subject or concept, followed by application of the algorithm and then iterative refinement by additional human input. Legal teams can then focus on priority categories to expedite review.

Many categorization technologies do more than just categorize; they also rank documents in terms of the strength of association within the cluster. This feature can also be used to speed up document review by allowing legal teams to concentrate only on the documents with the highest rankings.

Today, we have predictive coding.

The Grossman-Cormack TAR Glossary defines predictive coding as "[a]n industry-specific term generally used to describe a [TAR] process involving the use of a machine learning algorithm to distinguish relevant from non-relevant documents, based on subject matter expert(s)' coding of a training set of documents." The goal of predictive coding software is to assign prioritization scores to individual documents based on likelihood of relevance. Higher scores indicate that documents are more likely to be relevant while lower scores indicate that documents are less likely to be relevant.

The science underlying predictive coding algorithms is not new. In fact, it has been used for decades in other industries such as energy distribution, air traffic control, weather forecasting and insurance coverage, among others. Any field where known facts can be extrapolated and monitored with a statistically sound control model can successfully implement this science. The technology can be seen in everyday applications – whether it be online merchants that recommend products based on prior purchases like Amazon and Zappos, or entertainment sites such as Pandora or Netflix.

Predictive coding requires human involvement. The system first presents a randomly selected set of documents to subject matter experts (SMEs) for review and calculates an initial score based on yes/no determinations about the documents' relevance. This initial selection becomes the control set that will later be used for statistical verification. Then, SMEs review several smaller sets of documents that

are selected by the system to optimize its learning. During this latter phase, the predictive coding system includes some documents that were previously presented to the SMEs, as well as ones that are similar to previously selected documents, in order to ensure consistency. Behind the scenes, the system is generating a set of weighted attributes to include and a subset to exclude, which will result in optimal recall[1] and precision[2] outcomes. At a certain point, using indicators supplied by the predictive coding engine, the system is said to have stabilized.[3] Some predictive coding systems will then apply a relevance score to documents, resulting in a ranking from the most relevant to the least relevant. Other systems classify documents as "likely relevant", "likely not relevant" and "unclassified". At this point, legal teams and eDiscovery experts use this information to make decisions about documents that require additional review.

Ultimately, it is essential to remember that predictive coding systems are one TAR approach but not all TAR approaches are necessarily predictive coding.

Primary TAR categories: "Concept-Based" and "Support Vector-Based."

According to the Grossman-Cormack Glossary, TAR "is a process for prioritizing or coding a collection of documents using a computerized system that harnesses human judgments of one or more [SMEs] on a smaller set of documents and then extrapolates those judgments to the remaining document collection. Some TAR methods use machine learning algorithms to distinguish relevant from non-relevant documents, based on training examples coded… by the [SMEs], while other TAR methods derive systematic rules that emulate the expert(s)' decision-making process. TAR processes generally incorporate statistical models and/or sampling techniques to guide the process and to measure overall system effectiveness."

Thankfully, TAR users do not need to be statisticians who fully understand the theoretical minutia behind these applications, but it is helpful to understand the basics of the two primary approaches.

Concept-Based TAR

Concept-based TAR systems analyze the context of words within a set of documents and translate that information into mathematical models in order to build a conceptual search index. Once a model has been built, a "find more like these" algorithm is applied to the document population to identify documents that are similar in conceptual content.

Concept-based search engines are a clear improvement over keyword searching alone. These systems return more potentially content-relevant documents without the limitations of Boolean logic and false positives can be suppressed through document seeding.[4] The ranking that is obtained from the categorization process measures how closely the documents resemble the exemplars provided to the system. When trying to find documents that closely resemble each other, concept-based TAR technology is appropriate.

On the plus side, concept-based TAR applications are being integrated into review platforms and can easily be deployed if the use of advanced technology is warranted. Conversely, early use of concept-based technology outside of a review platform is nearly impossible, making it difficult to use this technology for culling. In addition, some experts criticize concept-based TAR because it tends to take longer to reach system stability than SVM technologies. This is largely because concept-based TAR requires a post-categorization "overturn correction workflow"[5] in order to reach the desired confidence level. System stability for concept-based technology is normally achieved after review of approximately 10,000 to 20,000 documents (fourteen to twenty one days of training, sometimes longer). Finally, with concept-based TAR, not all documents receive a score. Documents that do not fit into "relevant" or "not relevant" categories are tagged as "uncategorized."

Support Vector-Based TAR

TAR applications based on Support Vector Machine (SVM) technology, including predictive coding, are also well-established analytical modeling technologies. Most SVM based TAR applications automatically establish key benchmarks of recall and precision of a given document population early in the process.

Proponents of this approach appreciate being able to use support vector-based TAR to limit the volume of ESI that must be loaded for review based on prioritization of datasets. The scores and rankings applied by SVM algorithms can be exported for clients to use in the review application of their choice. Another clear advantage is that SVM systems typically take only three to five days, or a review of 2,500 to 3,500 documents, to train.

Use cases and workflows.

TAR applications can be used successfully at several stages of a case. The specific workflow chosen depends on case variables and the legal team's objectives. However, when appropriately documented, these workflows strengthen the results and defensibility of a party's review and production.

For example, support vector-based TAR can be used for early case assessment on datasets collected from important custodians in order to isolate key documents as part of a settlement risk analysis. SVM systems can also be used to identify an initial set of keywords and concepts to use in negotiations with opposing counsel when developing search term protocols. Furthermore, SVM systems can be utilized to rank documents in order of likelihood of relevance to establish review priority, especially in situations where tight timeframes constrict review of a large document population. On the other hand, concept-based TAR approaches can be used during deposition preparation to identify documents that relate to important issues of the case.

Finally, upon completion of review, a legal team can use TAR to conduct quality assurance checks by setting up a confusion matrix and comparing the relevance designations of the review team to the system's relevance scores (see table).

	Human review/ relevant	Human review/ not relevant	Totals
TAR/ relevant	(produce) 3,048	*2,531*	5,579
TAR/not relevant	*1,576*	(don't produce) 40,495	42,071
	4,624	43,026	47,650

In this example, there were 3,048 documents that the review team and the TAR application agreed were relevant and would be included in the production. There were 40,495 documents the review team and the TAR application agreed were non-relevant and would not be produced. The scope of the quality assurance effort was therefore focused on the two subsets for which the TAR system and the review team disagreed on the determination (the 2,531 documents that the review team marked as non-relevant but the TAR program scored as relevant and the 1,576 documents that the review team marked as relevant and the TAR program marked as non-relevant, indicated in green boxes above).

These 4,107 documents were submitted to senior reviewers for second pass review and verification. The senior team found that almost 1,500 additional documents were, in fact, responsive. The responsive set therefore increased from 4,624 to just over 6,000. The lead members of the legal team comfortably concluded that a defensible and reasonable effort had been made to identify responsive documents.

...a few seminal decisions provide guidance on the defensible use of TAR and the importance of agreed-upon protocols and disclosure – both of the processes utilized and the results achieved.

TAR case law.

At this time, there are very few cases that explicitly deal with TAR. The decisions highlighted below, taken chronologically and collectively, neither mandate nor prohibit the use of TAR technology. There is a trend towards judicial awareness of TAR and support of parties' agreements to use advanced technologies. Additionally, a few seminal decisions provide guidance on the defensible use of TAR and the importance of agreed-upon protocols and disclosure – both of the processes utilized and the results achieved.

- A court acceded to the defendant's request the use of TAR over the plaintiff's objection in ***Global Aerospace, Inc. v. Landow Aviation***[6]. The plaintiff argued that the court's approval of TAR constituted its agreement that Defendants need only produce a certain percentage of responsive documents. The court allowed the use of TAR, but also preserved the plaintiff's ability to challenge the completeness and sufficiency of the production.

- ***In Kleen Products LLC v. Packaging Corp. of America***[7], Plaintiffs sought to compel the defendants to redo their previous productions using predictive coding technology. Defendants countered that their use of keyword searching and advanced analytics (such as email threading) was precise and thorough. After a protracted struggle involving two separate hearings and multiple written submissions, Plaintiffs withdrew their demand as to the existing document request. For future production requests, the parties agreed to meet and confer regarding production methods, raising the possibility that Plaintiffs might renew their demands about TAR.

- The court issued an order directing both parties to demonstrate why they should not use a single predictive coding provider in ***EORHB, Inc. v. HOA Holdings, LLC***[8]. The court's decision in *EORHB* was especially bold at the time, because it essentially prescribed the use of TAR technology (and one provider thereof) unless the parties could show cause not to do so. However, after the show cause hearing, the court lifted the order. This allowed Defendants to proceed with the vendor of their choice and Plaintiffs to proceed without using predictive coding, given the low volume of documents they were expected to produce.

- ***In re Biomet***[9] was a products liability case in which the defendant began review by using search terms to cull nearly 20 million documents over the objection of Plaintiffs' Steering Committee. The Committee urged the defendant to await a ruling from the Judicial Panel on Multidistrict Litigation before proceeding with document productions. When Defendant thereafter used predictive coding, Plaintiffs argued that the initial use of search terms had tainted the document pool and asked the court to require the defendant to start over, using only predictive coding. The court refused to honor Plaintiffs' request on the grounds that Defendant had fulfilled discovery obligations using reasonable methods.

Although courts have not mandated or prohibited the use of TAR outright, they have specified acceptable rules of engagement for the use of TAR. Perhaps the most well-known case is ***DaSilva Moore v. Publicis Groupe***[10]. *DaSilva Moore* is touted as the first published judicial opinion that implicitly approved the use of TAR for reviewing documents. Both parties assented to the use of TAR, but argued over the manner in which it would be employed; per the parties' wishes, the court permitted TAR usage and provided guidance to establish rules of engagement. U.S. Magistrate Judge Andrew Peck emphasized that parties intending to use TAR must choose a reliable vendor and program and must design an "appropriate process" that includes "appropriate quality control testing."

The opinion specified a number of steps that parties should consider, including the following suggestions:

1. Bring both vendor experts to a court hearing to respond to the magistrate judge's questions.
2. Allow the requesting party to view the documents that were used to train the TAR system, both those that were marked responsive and those that were marked not responsive.
3. Allow the requesting party to view the additional documents that were used to stabilize the predictive coding system, whether or not they were marked responsive.

4. Do not adhere to an arbitrary number of documents that will be produced without reference to the statistical results.

5. Do not limit the number of iterative reviews used to "train" the system up front, but rather assess whether the system has stabilized before stopping the iterative reviews.

U.S. District Judge Andrew Carter later approved Judge Peck's protocol on the basis that it was reliable, standards-based and allowed participation by Plaintiffs, who were given the ability to object once production was made if they believed it to be insufficient.

The DaSilva Moore principles of cooperation, transparency and attention to a strong workflow had a concrete and defined role in the case of In re ***Actos***[11]. In Actos, the parties entered into a detailed protocol, wherein three experts from each side met, collaboratively reviewed the training set of Defendants' documents and agreed upon relevance determinations for all non-privileged documents in the set. Robust protections to guard documents subject to privilege and confidentiality were included. This approach may not be right for every case in which TAR is used, but it does provide an example of a transparent process.

Although there are few definitive decisions regarding Technology Assisted Review, there is certainly room for optimism that these applications will succeed in gaining widespread judicial support. As lawyers and judges begin to trust the reliability of TAR and especially predictive coding, they will fully realize and understand when and how to best deploy advanced technologies in different scenarios.

TAR and Defensibility.

Parties can choose, but cannot be forced, to use TAR applications to help them identify relevant documents and make decisions about review. Producing parties who use TAR can expect the receiving party to press for transparency, sharing of the training set, disclosure and perhaps agreement on key statistical metrics and possibly joint review of the training set to agree upon relevance, among other criteria. However, it is not yet known whether courts will require parties to share this information. On the one hand, courts traditionally have avoided micro-managing the details of a party's decisions about a production unless it is challenged as insufficient. On the other hand, a producing party who uses new and complex technology as a device to more efficiently identify relevant documents should expect the receiving party to ask questions about the methodology.

A defensible protocol is not characterized by particular statistical metrics or the number of sample documents reviewed, but rather is based on a documented workflow that details steps to be followed during the TAR process. Initial sampling and review, system stabilization and sampling of the results are all factors that should be considered. Some of these details will probably need to be shared with the receiving party, but the extent of sharing required is not yet clear.

TAR and Proportionality.

In addition to their obvious potential for analysis and culling, TAR technologies can be instruments of proportionality. With sound workflows and reliable sampling, parties can estimate early in the discovery process how much review must be conducted to achieve a production of the most relevant documents. With this knowledge, review costs can be calculated and informed decisions can be made about which documents to review. Moreover, parties can decide whether to request relief from the court, through discovery limitations or cost shifting.

Conclusion

In light of large data volumes, prohibitive eDiscovery costs and tight deadlines, Technology Assisted Review should always be a consideration. While not every case warrants the use of TAR – such as matters with less than 20,000 documents – the benefits of properly-used advanced technologies are undeniable. Parties are incorporating TAR workflows early in their cases to cull irrelevant documents so that more manageable, substantive datasets can be identified for final review and production. Both producing parties and receiving parties are finding their way with TAR, with relatively little guidance from courts. However, current and future users can be sure that detailed planning and careful attention to protocol will improve the results and defensibility of TAR.

Sources:

1. "Recall" asks "what percent of the relevant documents were retrieved by the algorithm?"
2. "Precision" asks "what percent of a given sub-set of documents are relevant?" and is inversely proportional to recall.
3. "Stabilization" occurs when the system has learned all it can in order to predict relevance.
4. "Document seeding" is presenting the system with documents that the Subject Matter Expert has already deemed relevant to better train the element.
5. "Overturn correction workflow" is a process that reverses the system's incorrect document classifications
6. Global Aerospace, Inc. v. Landow Aviation. No. CL 61040 (Va. Cir. Ct. Loudon County Apr. 23, 2012).
7. Kleen Products LLC v. Packaging Corp. of Am., No. 10 C 5711, 2012 WL 4498465 (N.D. Ill. Sept. 28, 2012).
8. EORHB, Inc. v. HOA Holdings, LLC, C.A. No. 7409-VCL (Del. Ch. Oct. 15, 2012).
9. In re Biomet M2a Magnum Hip Implant Prods. Liab. Litig., No. 3:12-MD-2391, 2013 WL 1729682 (N.D. Ind. Apr. 18, 2013).
10. DaSilva Moore v. Publicis Groupe, 11 Civ. 1279, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).
11. In re Actos (Pioglitazone) Prods. Liab. Litig., MDL No. 6:11-MD-2299 (W.D. La. July 27, 2012).

D4

D4 is the technology and eDiscovery division of Special Counsel, specializing in delivering exceptional, flexible and innovative eDiscovery solutions and managed services.

SPECIAL COUNSEL

At Special Counsel, we are a leading provider of legal consulting, attorney recruiting, legal talent, legal technology and eDiscovery solutions, with a growing international footprint. But there's more to being a leader than having tremendous size and scale. Our strength comes from our ability to partner with clients to create solutions on a case-by-case basis. It comes from our gift for fostering, furthering and promoting the expertise of our people. And it comes from our commitment to doing business with transparency, flexibility and integrity. That's what sets us apart. That's what makes us Special. To learn how we can help you, visit specialcounsel.com.

Special Counsel is a proud part of the world's leading HR solutions partner, the Adecco Group, serving as its legal solutions business unit.


D4