opentext™

WHITE PAPER

Choosing the right technology-assisted review protocol to meet objectives

When used correctly, TAR has the potential to offer tremendous savings, both in review time and cost, without sacrificing the quality of results

opentext™

Contents

Using any technology-assisted review (TAR) protocol will undoubtedly reduce the time and expense of reviewing electronic stored information (ESI) over traditional linear review. But, getting the best results will depend on carefully matching project objectives and constraints with the inherent strengths and weaknesses of predominant TAR techniques and, in some instances, combining TAR protocols. This white paper provides the necessary background and identifies the pertinent considerations to facilitate selection of the appropriate TAR protocol for typical use cases across the legal landscape.

1. What is technology-assisted review?

TAR, also known as predictive coding or computer assisted review, is a process whereby humans leverage technology to efficiently identify specific documents in a vast and disorganized corpus. Every TAR system encompasses human review for a portion of a document collection to train computers that, in turn, extrapolate those human judgments to the balance of the collection, enabling faster and more cost-effective review.

The Grossman-Cormack Glossary of Technology Assisted Review defines TAR as: "A process for prioritizing or coding a collection of documents using a computerized system that harnesses human judgments of one or more Subject Matter Expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining document collection."[1]

What exactly does this mean? Think of modern TAR systems as a music app for documents. A music app's goal is to find and play music that the listener likes, interspersing songs from favorite artists or genres with new songs that share key characteristics, known as "features." While the music app has millions of songs in its archive to choose from, it does not initially have any ability to guess what the listener wants to hear—until it learns to do so.

It learns by extrapolating from as little as a single artist, song or genre identified as a favorite. Based on that fairly generic starting point, it then begins to choose additional songs that have certain similarities. The reviewer provides what is known as relevance feedback, grading its selections by clicking a "thumbs up" or "thumbs down" button.

Based on this training, the app's algorithm analyzes a complex array of features, such as melody, harmony, rhythm, form, composition, style and vocalist, to differentiate the songs the reviewer likes from those he or she disliked. The more feedback the reviewer provides, the smarter the system gets. Eventually, a customized station will play mostly music the reviewer enjoys, with only an occasional miscalculation.

The modern TAR process works similarly. The TAR algorithm learns, from its human partner's feedback, which documents are relevant, with algorithmic judgments improving over a period of time. With TAR, a human reviews a document and tags it as relevant or not relevant. While other tags are possible for other applications, for simplicity this section only discusses relevance searches.

In the background, a computer algorithm continuously observes the assigned tags and uses that input, together with the features (typically, words and phrases) to make comparisons between the tagged documents and the remaining documents in its set. The algorithm then ranks every document in what it calculates to be the likelihood of relevance, shuffling documents that are most likely to be relevant (e.g., the highest ranked documents) to the top of the pile for human review, just as the music app shuffles the songs it expects the listener will enjoy to the top of the playlist.

This iterative process continues, cycling through review, analysis and ranking, until the review is discontinued. The objective of the review determines how long the process will continue, a decision that is made by the human review team, not the computer.

Of course, the objectives of TAR are considerably more serious than those of a music app, so review teams must consider a variety of options, techniques and strategies based on the goal.

When used correctly, TAR has the potential to offer tremendous savings, both in review time and cost, without sacrificing the quality of results. With TAR, review teams can work faster and process documents that are most likely to be relevant first. A relatively simple sampling process within TAR, showing the percentage of relevant documents found, can also give the review team a reasonable, defensible basis for concluding a review when the search objectives have been satisfied.

2. TAR protocols and the progression from TAR 1.0 to TAR 2.0

There are three basic TAR protocols. Simple passive learning (SPL) and simple active learning (SAL) are typically associated with early versions of TAR, now known as TAR 1.0. With simple learning, the algorithm is trained by a human reviewer until it develops a model of responsive documents that either stabilizes or reaches an acceptable level of quality. From that point on, the algorithm ceases learning and uses the information it gained in training to either classify or rank document sets.

SPL and SAL are differentiated by the set of documents they use for training. SPL typically uses randomly selected documents to train the algorithm. SAL usually starts with a set of clearly relevant and clearly non-relevant documents, often called the "seed set." From there, an SAL protocol actively selects the "gray area" documents in the collection for training, the ones that are most difficult to classify. This is called "uncertainty sampling." For both of these protocols, all training is completed by the SME at the beginning of the process, before review can begin in earnest. Once the algorithm stabilizes, training is complete, the review size is fixed and additional review is not required to improve the model.

A newer protocol, continuous active learning (CAL), is central to the second generation of TAR protocols, known as TAR 2.0. With CAL, the algorithm learns and improves continuously throughout the review process. Instead of a preliminary training phase, the human review team simply begins review while the algorithm observes those decisions and adjusts its criteria for determining relevance. Every review decision, from the first to the last, is used to train and improve the algorithm, ensuring that the most likely relevant documents are being ranked toward the top of the list, so they can be preferentially made available to reviewers.

The market has largely shifted toward adopting TAR 2.0 due to a variety of advantages. In particular, CAL has been shown to reach higher levels of recall, identifying a greater number of relevant documents more quickly and with less human review effort than either of the TAR 1.0 methodologies.[2] This allows organizations to meet tight production timelines, leverage a limited staff of human reviewers and minimize the bottleneck caused by the algorithm training process. CAL can also readily accommodate both changes in the scope of discovery and rolling data productions, since it continues training throughout the life of the review process. Its benefits have inspired CAL applications that extend beyond outbound productions, as discussed below.

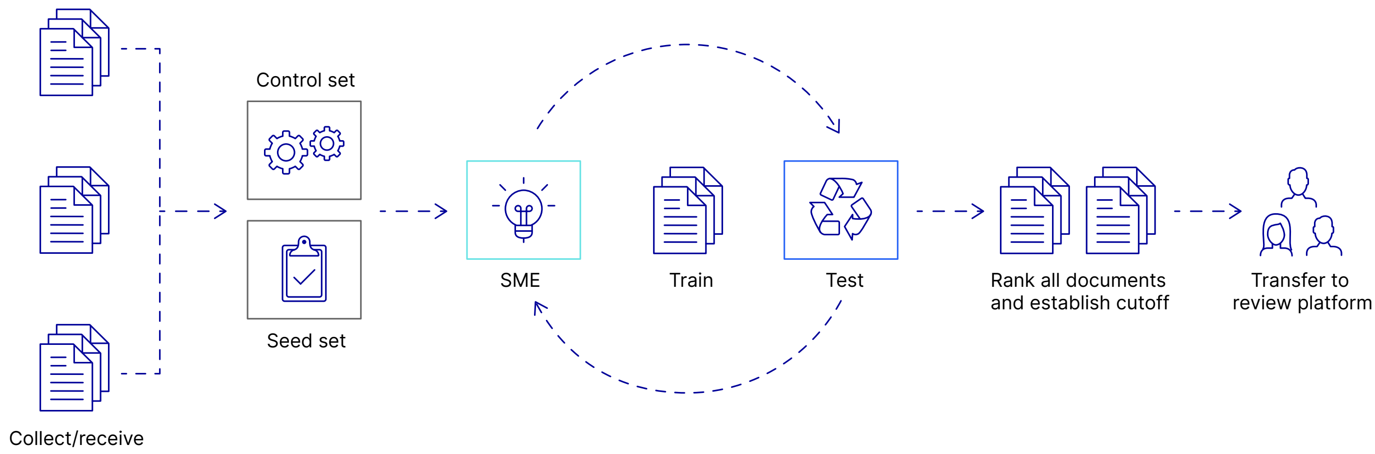
But the rise of TAR 2.0 does not spell the end of TAR 1.0, nor does it eliminate combining aspects of both protocols to achieve certain goals. Determining which protocol may be the best fit for a particular matter depends on objectives and requires a more detailed understanding of the various methodologies and preferred use cases.

Both TAR 1.0 and TAR 2.0 operate through an iterative cycle of reviewing documents, analyzing the results and managing the remaining documents. But, there are a number of specific differences, all of which stem from one critical distinction. A TAR 1.0 algorithm stops training when it stabilizes, regardless of how many documents are subsequently reviewed, whereas a TAR 2.0 algorithm is trained by every coding decision until the review stops. As a side note, the reader may see reference to future generations of TAR, such as TAR 3.0 or even predictive coding 4.0 systems, but they actually fall under the TAR 2.0 ambit. They are all based on a CAL protocol, discussed below, and modified to accommodate different training techniques. Neither is discussed in this white paper.

This white paper will next take a closer look at the workflows for TAR 1.0 and TAR 2.0.

3. TAR 1.0: One-time training

Figure 1 below is a diagram of a typical TAR 1.0 process, from the collection of the document set through the final review.



This is how a typical TAR 1.0 process works:

1. **Collection.** The first step in the protocol is to amass and process the entire collection of documents subject to review. From the TAR perspective, processing entails breaking each document into features (most often words or phrases) that will be used by the TAR algorithm to compare and rank or classify the documents for review purposes. And, as discussed below, because most TAR 1.0 systems depend upon a control set, it is critical to amass the entire collection before review begins. Otherwise, it may be necessary to re-initiate the entire TAR 1.0 process, particularly when new documents addressing new concepts are added to the collection, such as engineering documents added to a collection of primarily sales documents.

2. **Control set.** The next step in the protocol is to draw a random sample, typically 500 or more documents, that will be set aside and used as a control set to monitor progress and will not be used to train the algorithm. Before anything else can be done, the control set needs to be reviewed and coded by a subject matter expert (SME), usually a senior lawyer on the case. It is particularly important to have an SME review the control set, because it operates as the answer key or "gold standard" against which the algorithmic model will be compared to evaluate progress throughout the TAR process. This means it needs to correctly reflect the appropriate notions of relevance. And, to be effective, the control set must be representative of the entire collection of documents being reviewed, which is why the collection needs to be complete at the outset.

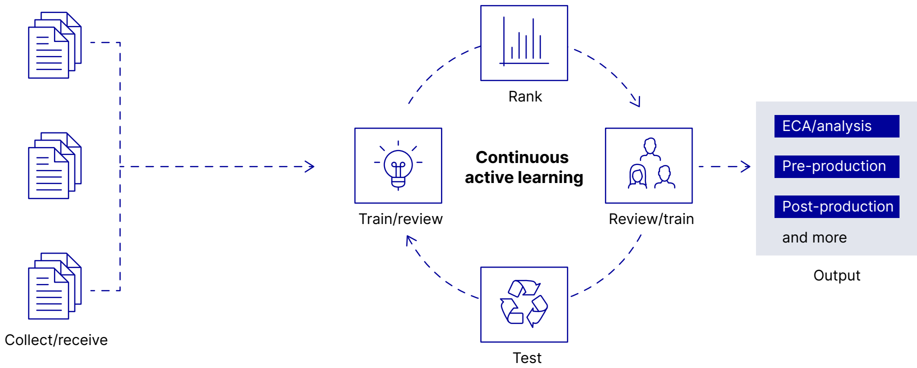
3. **Seed set.** The need for a seed set in a TAR 1.0 process depends upon whether it follows an SAL or SPL protocol. As an SPL protocol depends only upon randomly-selected documents to train the algorithm, there is no need for a seed set to initiate training. SPL, on the other hand, uses uncertainty sampling techniques to identify appropriate training documents. Before an SPL algorithm can find that uncertainty boundary, it needs to have some idea of what is considered relevant and what is considered non-relevant. That information comes from the review and coding of a seed set that provides good examples of both relevant and non-relevant documents. Typical SAL algorithms perform better, with roughly 50 relevant and 50 non-relevant examples in the seed set. As with the control set, the seed set needs to be coded by an SME to ensure accurate decisions and, in turn, appropriate selection of training documents.
4. **Training.** Once the control set, and perhaps the seed set, have been reviewed and coded, the SME continues the training process by reviewing batches of documents selected by the TAR engine, either randomly (SPL) or through uncertainty sampling (SAL). Each document is tagged as relevant or non-relevant. The training rounds typically involve review of between 1,500 and 5,000 documents. This training takes time. Assuming a reasonable review rate of 60 documents per hour, it will likely take the SME more than 65 hours just to stabilize the algorithm before review can start in earnest.
5. **Ranking and testing.** Periodically throughout the training process, the TAR algorithm analyzes the SME's tags and modifies and improves its relevance model. The algorithm typically tests the model by applying it to the documents in the control set to see how well it matched the SME's judgments.
6. **Stability.** Training, ranking or classification and testing continue until the algorithm's model is "stable." That means it no longer improves identifying relevant documents in the control set. For example, say the model correctly identified 75 of the 87 relevant documents in the control set. Over a few more rounds of training, the results do not improve, which generally means that, even with additional training, the algorithm will not get any better at finding relevant documents in the control set and, presumably, will be as good as possible when applied to the collection.
7. **Rank or classify the remaining documents.** When training is complete, the next step is to run the model against the entire document population. Doing so can take several hours depending on the system, or it may need to run overnight. This is a one-time ranking or classification based on SME training. Once the algorithm finishes ranking or classifying the collection, the algorithm is not given any more documents for training and can no longer improve based on further tagging by the review team.
8. **Generate and validate the presumptively relevant set.** Once the algorithm is applied to the entire collection, it will be split into two subsets: one that is presumptively relevant and one that is presumptively non-relevant. The documents that are presumptively non-relevant, called the null set, will generally be discarded and will not be reviewed any further. The presumptively relevant set may or may not be reviewed, as discussed below. There are two predominant methods for checking to validate the presumptively relevant set, ensuring that it has a sufficient number of responsive documents to meet any recall objectives. Often, the control set is used to set a cutoff. For example, if the user wanted to produce 80 percent of the relevant documents, they must find the rank in the control set where 80 percent of the relevant documents were located and simply produce everything above that rank. Otherwise, and particularly for classification algorithms, the user can take a random sample of both the presumptively relevant set and null set and determine the fraction of the total number of relevant documents found.

opentext™

9. **Conduct the review.** Once complete, the review team may be directed to look at the presumptively relevant documents or decide to produce those documents without further review. The user can also do a prioritized review, where the team looks at all of the documents collected based on their relevance ranking. That accomplishes two goals. First, if relevant documents are pushed to the top of the ranking, the team will see documents that are more likely to be relevant first. Second, once the team runs out of relevant documents, it can move quickly through the nonrelevant ones without fear of missing something important.

4. TAR 2.0: Continuous active learning

As pointed out in Figure 2 below, continuous active learning (CAL) is the hallmark of a TAR 2.0 protocol. A CAL system continually learns as the review progresses and regularly re-ranks the document population based on what it has learned to move the most likely relevant documents to the top. As a result, the algorithm gets smarter and the team reaches its goal sooner, reviewing fewer documents than would otherwise be the case with one-time training.



Here is how the TAR 2.0 protocol works:

1. **Collection.** As with TAR 1.0, the first step in the TAR 2.0 protocol is to amass and process a collection of documents, making the features of the documents available to the TAR algorithm. However, because CAL continuously ranks the entire document collection and training takes place throughout the review, it is not necessary to gather the entire collection before review begins. Engineering documents will simply be folded into the collection of sales documents and ranked based on the features of every document coded to that point in time. And, if they are relevant, the engineering documents will eventually be ranked near the top of the list and come up for review in due course.

2. **No control set required.** A control set is not necessary and not used in a TAR 2.0 protocol.

3. **Initial seeding**. The user can initiate a TAR 2.0 protocol with as many, or as few, documents as desired. One of the best ways to initiate ranking is to start by finding as many relevant documents as possible and feed them to the system to help train the algorithm or create a synthetic document to use as an initial seed. The user can even begin without any seed documents—just start reviewing, and the algorithm will learn based on every relevant and non-relevant document the user codes. Random sampling is generally not recommended for the purpose of initial training, since it is it is not necessarily an efficient means of finding relevant documents and is particularly problematic for low richness collections.

opentext™

4. **Begin review.** The review team can start immediately; there is no need for a subject matter expert to review any documents whatsoever. Reviewers will quickly begin seeing batches containing mostly relevant documents.
5. **Quality control.** As the review progresses, the subject matter expert, such as the senior attorney, can cull a small percentage of the documents to ensure that the reviewers are aligned with the proper scope of relevance. An effective TAR system will include a quality control algorithm that locates and presents those documents that are most likely tagged incorrectly.
6. **Finish.** The user continues until the desired recall rate is reached. They can track progress as the review progresses to see when it is time to stop.

The user can demonstrate success through a random sample of the unseen documents, called an "elusion sample." It will show how many relevant documents the user may have missed, from which recall can be calculated, as well as where one is in the review and, where appropriate, how many more documents are needed to reach the goal.

The process is flexible. Users can start with as many training seeds as they like or create a synthetic document. After the initial ranking, the team can get going on the review. As they complete batches, the ranking engine takes their new judgments into account and keeps getting smarter.

5. Key differences between TAR 1.0 and TAR 2.0

The TAR 1.0 process comes with a number of practical problems that limits its effectiveness.

TAR 1.0 requires SMEs for training. With TAR 1.0, to make certain that the algorithm correctly reflects the pertinent characteristics, an SME must review thousands of documents before the algorithm is ready to use. This creates a bottleneck where discovery cannot proceed until the SME—whose time is likely both limited and expensive—has spent 60 or more hours training the TAR system.

The SME must train the TAR 1.0 algorithm until it stabilizes. In document collections with low richness or numerous distinct issues, this may require the SME to review thousands of additional documents.

You only get "one bite at the apple." The better the algorithm is trained, the more accurately it can identify the most likely relevant documents and classify them, or rank them highly, so they ultimately end up in the presumptively relevant set. Yet, with TAR 1.0, the training period is limited, so the algorithm cannot incorporate additional feedback or continue to improve. Because the bulk of review occurs after the TAR 1.0 algorithm has evaluated and segregated the collection, lessons learned during review cannot inform the algorithm's operation.

TAR 1.0 is not flexible or adaptive. Because the TAR 1.0 algorithm is fully trained before review begins, it does not accommodate changes to the scope of discovery that occur during review, such as the addition of documents in a rolling production. It also cannot adapt to an evolving understanding of the case or the legal issues involved.

The structure of TAR 1.0 invites legal challenges. Opponents may object to the seed set or the protocol that was used to train the algorithm. Because training is a limited process, weaknesses in its foundation are both readily apparent and potentially fatal.

opentext™

TAR 2.0 solves many of these challenges

Rather than using an SME, a control set or a seed set, with CAL the human review team simply begins review while the algorithm learns in the background, analyzing tags and developing its sense of which documents may be relevant. The more comprehensive the initial relevant documents, the faster the algorithm will learn. The algorithm continuously ranks the entire document set, feeding more and more relevant documents to the human review team and continuously learns, adjusting and adapting throughout the entire review process.

TAR 2.0 eliminates the bottleneck caused by SME training, freeing senior attorneys to focus on finding relevant documents for training using other analytics and running quality control checks. CAL's continuous ranking and learning also obviate the problems of low data richness, rolling productions and changes to the scope of discovery. Finally, because there is no defined seed set to evaluate, TAR 2.0 minimizes the degree to which the algorithm training protocol can be challenged. Instead, all review is training and all training is review.

There is a characteristic dichotomy in the operation of both TAR protocols. A TAR 1.0 protocol will "train" more quickly than a TAR 2.0 protocol. In other words, if the user did the initial training of a TAR 1.0 algorithm and a TAR 2.0 algorithm with the same fixed number of documents and then just quit training, it would be necessary to review more documents using the TAR 2.0 ranking than the TAR 1.0 ranking. However, since a TAR 2.0 system never really stops training, it will eventually be more efficient than a TAR 1.0 algorithm. So, while TAR 1.0 may train more quickly than TAR 2.0, TAR 2.0 will ultimately be more efficient.

6. Choosing the right protocol: Start with the end goal in mind

Given this dichotomy, when starting a TAR review, ask: "Do I intend to review every document that will be produced?" If the answer is yes, the review team will review fewer documents using a TAR 2.0 protocol. If the answer is no, the team will review fewer documents using a TAR 1.0 protocol. Indeed, the user will only review the control set, possibly the seed set and enough documents to train the algorithm to stability, but undoubtedly more non-responsive documents will be produced.

7. When should TAR 2.0 be used?

The short answer is that TAR 2.0 should be used for the majority of review and production tasks. The success of any review can be measured by balancing the recall or completeness of the results (the percentage of relevant documents identified) with their precision or purity (the percentage of retrieved documents that are actually relevant). Typically, a technique that increases one of these metrics will decrease the other, so it helps to be explicit about the goals of review from the outset. TAR 2.0 using CAL can be deployed rapidly to maximize either recall or precision, which makes it amenable to a wide range of use cases.

Classification tasks

The most familiar application of TAR is still the classification of an outbound production for eDiscovery. While this review should strive for reasonably high percentages of recall and precision, it is guided by the principles of reasonableness and proportionality, not perfection. Recall is valued more highly than precision, but a modest target of 80 percent recall is a common standard, enabling the cost and effort of the search to remain proportionate to the value of the case. TAR 2.0 excels in classification tasks due to the rapid results it produces and its ability to immediately focus human review efforts on the relevant documents. As mentioned above, CAL is particularly useful in outbound production instances where there will be rolling uploads, or where the scope of discovery is anticipated to be complex or evolving.

Knowledge generation tasks

In investigations, early case assessment (ECA) and review of received production sets, the goal is knowledge generation rather than classification and time is of the essence. Knowledge generation tasks seek the best documents, such as those with the most interesting content and essential portions of the story. Precision is therefore key, while recall is relatively unimportant. In other words, users want to avoid reviewing documents that are not relevant wherever possible. Because TAR 2.0 skips the laborious and time-consuming training process or, more accurately, subsumes training into review, TAR 2.0 starts providing insights much more quickly than TAR 1.0. This allows reviewers to discover useful information and start discerning stories within the data almost immediately. By comparison, a TAR 1.0 SAL protocol focuses almost exclusively on documents for which relevance is uncertain, limiting the number of truly relevant documents that are available for review until the system has been fully trained.

One note though, all TAR methodologies rank documents according to how likely they are to be relevant, not how inherently interesting they are. While an unusual or atypical document may be highly valuable for reconstructing the story of the case, it might not be recognized by a TAR algorithm because it is so different from anything else that has been reviewed. That is especially true for a TAR protocol that relies on simple learning. Consequently, TAR 2.0 vastly outperforms TAR 1.0 for knowledge generation tasks.

Investigations present additional challenges that overwhelm TAR 1.0's capabilities. Unlike litigation, there are no fact-laden complaints to focus an investigative search and no broad seed sets to aid in training. This paucity of exemplars is not a problem for CAL algorithms, which can begin locating the majority of pertinent documents based on a single positive seed document. That single document may even be a synthetic seed, such as a recitation of known facts or a string of keywords generated to reflect the key language and concepts sought in the search.

When reviewing opposing party productions, the objective is to weed through the collection efficiently to identify particularly relevant documents. Rapidly surfacing those "hot" documents is another precision-oriented task that suits a CAL algorithm.[3] Since CAL can be initiated from a synthetic seed laced with the most critical details of the information that is being sought from the opposing party production, a TAR 2.0 algorithm will quickly recognize the features that make a document "hot" and elevate those for prompt review. And, if additional issues are discovered during review, the TAR 2.0 algorithm, unlike a 1.0 system, will seamlessly incorporate those into its relevance calculation.

Protection tasks

By contrast to other document review tasks, the objective of protection tasks is the absolute identification and protection from disclosure of certain types of information, such as privilege, trade secrets or confidential information. This essentially demands 100 percent recall, without exception, while precision is less important. The ability to rapidly develop a CAL algorithm using privileged documents makes TAR 2.0 an excellent option for conducting a privilege review when the bulk of documents will be produced without eyes-on review, as in a second request or a subpoena response. The TAR 2.0 review will quickly elevate likely privileged documents for withholding and the privilege review can cease once it appears that the algorithm is not finding any additional privileged documents.

However, the best way to maximize recall, e.g., to protect sensitive documents from disclosure, is to stack different techniques, rather than relying on one technique, as each methodology is prone to its own type of mistakes. Human reviewers, for example, tend to make random mistakes on individual documents, while TAR systems often make systematic errors, getting entire classifications of documents right or wrong. Combining different approaches, by layering TAR 2.0 with other review methodologies, such as keyword searching, eliminates the gaps inherent in each approach.

8. When should TAR 1.0 be used?

Although TAR 2.0 is the more efficient and appropriate solution for many review scenarios, it should not be the only review tool in an organization's arsenal. There are times when a TAR 1.0 approach, or a hybrid approach that combines the benefits of both SAL and CAL (discussed in the next section), is either preferable or essentially mandated. Organizations subject to a variety of eDiscovery obligations for a wide range of requesting parties should keep multiple options at the ready.

In classification tasks requiring outbound production, TAR 1.0 works particularly well for reasonable, cost-effective efforts to rapidly identify and produce requested documents without requiring a high level of recall or precision or the need to review all of the documents being produced. Because TAR 1.0 trains more quickly than TAR 2.0, TAR 1.0 is both cost-effective and efficient when the primary consideration is not technical perfection, but rather compliance with an affirmative duty to make reasonable efforts to find and produce requested documents.

The typical scenarios for which TAR 1.0 might be particularly useful include Hart-Scott-Rodino Second Requests and third-party subpoenas. When responding to a government second request, the documents sets are typically massive, with broad responsiveness criteria and very tight deadlines that make reviewing the entire production set impractical. In the third-party subpoena context, cost is the primary consideration, which means the review team wants to minimize the number of documents that need to be reviewed. And there is typically little interest in the true substance of the documents, since they are being produced in a litigation to which they are not a party. In both situations, the review team will often not be able to, or even want to, examine every document. The ability to train a TAR 1.0 algorithm quickly and inexpensively and generate a reasonable production set makes TAR 1.0 a particularly suitable alternative.

There are also situations where using TAR 1.0 may be inevitable, because of reviewing party demands and the inability to effectively negotiate any alternatives. For example, some regulatory and government agencies essentially "mandate" a TAR 1.0 approach, much like the approach taken by the Department of Justice in its published review protocol.[4] Typically, in those scenarios, the success of the TAR review is tied to TAR 1.0 statistics, making it difficult to comply with a TAR 2.0 approach. Additionally, although courts should not generally be involved in designing review methodologies, some courts, as well as counsel, lack familiarity with CAL, which has the tendency to focus the discussion of protocols on a TAR 1.0 approach.

9. Adapting to conditions: Combining aspects of TAR 1.0 and TAR 2.0

In some cases, legal teams need to remain flexible during a review project and pivot workflows between TAR 1.0 and TAR 2.0 approaches. Circumstances might include scope changes, risk tolerance and changing deadlines. Alternatively, users may want to use elements of the TAR 1.0 workflow in a review tool that does not have it built in. For example, a legal team may have originally believed that there was a fairly low risk in producing documents without human review and, as such, implemented a TAR 1.0 workflow. After some review has taken place, the risk assessment may have changed based on the documents found to date. With the flexibility to switch to a TAR 2.0 workflow, all work product done to date can be used without restarting the entire process. Conversely, if a review team started a TAR 2.0 workflow based on the need to review each document being produced, but timelines have changed to the extent that it is no longer feasible, using elements of a TAR 1.0 workflow could make sense. This may be referred to as a TAR 1.5 workflow.

10. Additional considerations when choosing a TAR methodology

Protocol negotiation. As mentioned above, a party may not always be in a position to negotiate the TAR protocol used in a particular production. Depending on the opponent or regulatory authority making the request and the strength of the user's position, a legal team may have to accede to a protocol that has been largely determined by someone else.

Transparency obligations. Be mindful that the TAR protocol chosen—and the degree of transparency about how it is being used—could expose an organization to unexpected risk. Especially with TAR 1.0, when operating details of the training and search protocols are shared with an opponent or court, any later adjustment or divergence that the user makes from that plan might trigger a legal challenge. Clear and accurate communication is always appropriate in eDiscovery, but users should not feel compelled to overexplain their TAR methodology. After all, review processes were largely confidential in traditional paper-based discovery—an approach that some courts have held is worth emulating with TAR.[5]

11. Conclusion

Vendors and proponents of TAR technology tend to present the choice between TAR 1.0 and TAR 2.0 as a binary decision that must be made wholesale, across all cases, using distinct eDiscovery systems. This is no longer true. Legal teams today can adopt bespoke methodologies and workflows that offer both TAR 1.0 and TAR 2.0 , even within the same eDiscovery platform platform.

Finally, whatever approach is chosen, remember that under both the ABA Model Rules of Professional Conduct and most states' ethical rules, attorneys have an ethical obligation to understand the risks and the benefits of all relevant technology.[6] By improving one's understanding of the various TAR methodologies and their suitable use cases, there is confidence in knowing ethical obligations in the process have been satisfied.

OpenText™ provides proprietary technology-assisted review technology, best practices guidance and support for each client project. The OpenText Insight eDiscovery and investigations platform offers both Insight Predict, TAR 2.0 based on CAL, and Cut Point Review, TAR 1.0, within the same UI for ease of use, validation and production. OpenText Axcelerate includes TAR 2.0 based on continuous machine learning, with workflows that can incorporate both TAR 1.0 and 2.0 based on a client's objectives.

OpenText also provides supporting end-to-end managed document review services leveraging technology-assisted review for the most expedient, accurate and cost-effective review possible.

opentext™

Sources:

1. *The Grossman-Cormack Glossary of Technology Assisted Review* (Fed. Courts L. Rev. 2013, Vol. 7, Issue 1).

2. Gordon V. Cormack and Maura R. Grossman, "Evaluation of machine-learning protocols for technology-assisted review in electronic discovery" (2014).

3. OpenText Insight Predict case study, "Using TAR to Find Hot Docs for Depositions."

4. U.S. Department of Justice, Antitrust Division, Predictive Coding Model Agreement.

5. Dynamo Holdings Ltd. P'ship v. Comm'r of Internal Revenue, Nos. 2685-11, 8393-12, 143 T.C. No. 9 (Tax Sept. 17, 2014) (noting that "the Court is not normally in the business of dictating to parties the process that they should use when responding to discovery. If our focus were on paper discovery, we would not (for example) be dictating to a party the manner in which it should review documents for responsiveness or privilege").

6. American Bar Association, "Rule 1.1 Competence - Comment."

About OpenText

OpenText, The Information Company, enables organizations to gain insight through market leading information management solutions, on-premises or in the cloud. For more information about OpenText (NASDAQ: OTEX, TSX: OTEX) visit: opentext.com.

Connect with us:

- OpenText CEO Mark Barrenechea's blog
- Twitter | LinkedIn

opentext.com/contact