CLOUD COMPUTING

LITIGATION AND PRACTICE SUPPORT
BEYOND THE BASICS
MANAGING BIG DATA

MAY | 13
DIGITAL
WHITE PAPER
i/+∧

Technology-Assisted Review 2.0

by Ignatius Grande of Hughes Hubbard & Reed LLP
and Andrew Paredes of Epiq Systems

Legal teams and their outside counsel must deal with an ever-expanding universe of electronic data. Traditionally, counsel reviewed every document in a case in a linear fashion, making use of search terms when possible. In combination with the growing volume of electronic data, this type of review process has proven costly and consumes valuable time during discovery.

Technology-assisted review (TAR), also known as predictive coding, is being used more commonly to help case teams bypass or modify linear review processes, to prioritize their review by relevance, and to cut costs and resource requirements while still meeting deadlines and production requirements in a defensible process.

TAR can be a powerful tool when applied correctly. Surrounding the technology with the right workflows, processes and quality control procedures is essential to ensuring quality output.

HOW TAR WORKS

Although different vendors conduct technology-assisted review with subtle variations, the steps of TAR are the same.

The first step is to get a representative sample of your set of electronic documents. This "seed set" can be created in different ways, but the most common is to use a random sampling of the full set of documents. Search terms can be used to help create the seed set; however, this method should be employed sparingly.

The next step involves having an attorney or team of attorneys who are knowledgeable about the subject matter of the case review the seed set and code each document for responsiveness and for any other attributes relevant to the case. Based on this initial coding, the TAR tool will begin to learn how to score documents for relevance.

The TAR tool will use this knowledge to create a new smaller set of documents for a senior attorney to review for

responsiveness. The results of that review will be put into the tool to allow it to continue perfecting its algorithm. Via an iterative process, the tool will keep feeding additional sets of documents to the attorney(s) for review. In doing this, the attorney(s) will "train" the algorithm by evaluating where their decisions differ from the computer's and making appropriate adjustments. This process will be repeated until the attorney is satisfied the tool is pulling responsive documents efficiently. After each iteration, the tool will provide statistical summaries of how successful the algorithm is in focusing on and ranking the responsive documents. The attorney team will work with the vendor using these statistics and their experience going through the document sets to determine when the algorithm has been perfected.

Although TAR is a powerful tool, it does not replace attorney review. TAR prioritizes the documents most likely to be relevant, and the vendor — in conjunction with the case team — will implement a process that will often include some attorney review before documents are produced in a case.

Technology-assisted review can be applied differently to each case. By using TAR, legal teams can opt to forego a linear or human responsiveness review entirely if their comfort level and the sensitivities of the review allow for it. Another frequent application is to have reviewers evaluate only those documents selected for production as an added safety check. In addition, statistical sampling often is included as part of the TAR process.

WATCH A RECORDING OF EPIQ'S "TECHNOLOGY-ASSISTED REVIEW 102: BEYOND THE BASICS" WEBINAR (SELECT PLAYBACK OPTION)

VIEW VIDEO

http://bit.ly/ZlTDJ8

BENEFITS OF TAR

TAR can be deployed quickly for comparatively low-value operations, such as looking for confidential or personal information. There are many more benefits to utilizing the power of technology-assisted review.

Prioritization: One of the primary benefits of TAR is the prioritization of the documents at hand. The algorithm feeds responsive documents to the legal team, placing the highest scores on documents deemed most likely to be responsive. There are different ways to deal with the output of documents. One method involves having an attorney team review this rich set of highly responsive documents. The remaining documents, or a subset of the remaining documents, (those deemed least likely to be responsive) can be sent to an offshore or lower-cost review team.

Alternatively, counsel can choose to run quality control and sample the "unlikely to be responsive" document set to ensure further human review is not necessary.

Quality of Review: Because TAR often can give you a confidence score of the responsiveness of each document, the process is more transparent than a traditional document review. The statistically based approach employed by TAR also lends itself well to the creation of useful reports. By looking at many data points, counsel can obtain information concerning the shared characteristics of groups of documents — information that may have gone unnoticed otherwise.

Turnaround Time: When you run into an eleventh-hour data problem (for example, someone discovers a cache of additional data or a network share that hadn't been previously collected), it's possible — depending on how different the new data set is from the one on which you've trained your model — to deploy the model and classify documents on a tight deadline. You can even process a large corpus of data on a short turnaround.

Privilege Classification: TAR can also be used to assist in classifying potentially privileged documents with a fairly high degree of recall. This can be helpful, especially in situations when you have a stipulation under FRE 502(d), which enables you to claw back any privileged documents that may be produced.

THE JUDICIARY WEIGHS IN: REAL-WORLD CASE LAW

Within the last few years, there have been several judicial opinions focused on technology-assisted review. In addition, statutes, rules and case law have made it clear lawyers must understand technology, including TAR and when it should be a consideration in a case.

In 2012, the ABA passed an amendment to Comment 6 to Model Rule 1.1, which states: "to maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with the technology." Further, the ABA Commission on Ethics noted the critical importance of this Comment given the growing importance of technology in modern practice. Whatever their role in a case, attorneys have an affirmative duty to understand how technology affects the case, whether TAR may be used and, if it is used, that it is used correctly.

The first judicial opinion in which technology-assisted review was discussed in depth was *Da Silva Moore v. Publicis Groupe*. In this opinion, Magistrate Judge Peck held that TAR "is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review." Up until this opinion was issued, no court had ruled on the use of TAR in a matter. Magistrate Judge Peck also emphasized the importance of using an appropriate process when deploying TAR:

> *"As with keywords or any other technological solution to e-discovery, counsel must design an appropriate process, including the use of available technology with appropriate quality control testing, to review and produce relevant ESI."*

Other notable judicial opinions involving technology-assisted review issued since *Da Silva Moore* include:

- *Global Aerospace v. Landow Aviation*, No. CL 61040 (Vir. Cir. Ct. Apr. 23, 2012): In April 2012, this Virginia Circuit Court ordered that defendants could use predictive coding despite the plaintiff's objections it wouldn't be as effective as manual review. The court ordered TAR be used and stated the receiving party would have an opportunity to question the completeness of the production at a later date.

- *Re: Actos (Pioglitazone) Products Liability Litigation*, No. 6:11-md-2299 (W.D. La. July 27, 2012): The court in this case issued a detailed case management order that supported the use of TAR and set forth specific requirements for establishing an initial sampling set, training the TAR algorithm and setting a responsiveness threshold.

- *EORHB, Inc., et al v. HOA Holdings, LLC*, C.A. No. 7409-VCL (Del. Ch. Oct. 15, 2012): In this case, neither party had considered using technology-assisted review, but the court ordered both parties to do so or show cause why they should not. Further, the judge encouraged both parties to use a single discovery vendor.

- *National Day Laborer Organizing Network et al. v. United States Immigration and Customs Enforcement Agency, et al.*, 2012 U.S. Dist. Lexis 97863 (S.D.N.Y., July 13, 2012) (Hon. Shira Scheindlin): Judge Scheindlin opined TAR techniques are more accurate than keyword searching.

  "Simple keyword searching is often not enough… parties can (and frequently should) rely on latent semantic indexing, statistical probability models, and machine learning tools to find responsive documents…these methods (known as 'computer-assisted' or 'predictive' coding) allow humans to teach computers what documents are and are not responsive…and they can significantly increase the effectiveness and efficiency of searches."

- *Re: Biomet M2a Magnum Hip Implant Products Liability Litigation*, 3:12-md-2391-RLM-CAN (N.D.Ind. April 18, 2013): In this case, the plaintiffs objected to Biomet's application of TAR, which included keyword searching to reduce the initial volume of information, followed by the use of TAR. The plaintiffs sought to require Biomet to start its TAR process from scratch with plaintiffs' input. The court, citing proportionality principles, concluded Biomet's efforts complied with discovery obligations under the civil rules.

As made evident by recent case law, judges are starting to expect attorneys will consider and at least occasionally use technology-assisted review. Accordingly, attorneys who have no understanding of TAR or who fail to consider TAR as an option may not be well-received in court.

WHEN AND WHERE TO USE TAR

Each case will differ in its goals, stakeholders, factual complexity, languages used, issues presented and levels of scrutiny. How do you determine when a case is well-suited for this technology? There are a few general best practices you can follow:

- **Large Sets of Documents:** TAR can be a good fit when you need to cull down a large data set in a short amount of time. TAR is used often in connection with second requests by the Federal Trade Commission and the Antitrust Division of the U.S. Justice Department that fit that model.

- **Internal Investigations:** Internal investigations lend themselves well to the use of TAR due to the absence of an opposing counsel or court applying strict oversight to the process. The primary advantages of TAR are its efficiency and prioritization. When conducting internal investigations, TAR allows a case team to focus on key documents quickly.

- **Litigation:** When using TAR in the context of litigation, the case team must determine whether they will disclose to the other side up front that they intend to use TAR. When using TAR, they also need to decide how much information about their process they want to share with opposing counsel. While case law is still developing in this area, federal rules and local rules regarding Rule 26(f) meet and confer conferences help guide parties in their communications with opposing counsel.

- **Tight Deadlines:** The short amount of time necessary to point a team to the most responsive documents is one of the most important benefits of TAR.

- **Issue Considerations:** Algorithmic review works best with large, relative determinations or a handful of discrete issues, rather than many unrelated issues.

RESOURCE SELECTION

TAR is a tool that can be used to great advantage in a case, but it is not a substitute for expertise and a carefully vetted process. TAR will absorb the knowledge an attorney has of case documents and will execute instructions based on what the attorney tells it to do. If counsel does not communicate effectively with the TAR tool or too many attorneys train the tool, the output will not be as accurate as possible, no matter how powerful the tool. The assistance of a good vendor or consulting team can be invaluable in maximizing the potential of your TAR tool.

Whatever an attorney's role in a case, it's important he/she be familiar and comfortable with the way in which TAR is being deployed.

- **Process**: Be sure there is a well-documented process in place for the application of TAR and that it can be adapted

to the matter at hand. Work with a vendor to develop this process before the initial review commences.

- **Early Use of Search Terms:** It is an open question as to whether it is appropriate to use search terms prior to using TAR (see *Re: Biomet M2a Magnum Hip Implant Products Liability Litigation*).

- **Expertise**: In a field where start-ups are prevalent and competition is fierce, experience counts. Consider both the amount of time your vendor has been deploying its TAR methodology and the expertise and background of its consultants and project managers.

- **References**: Prior to working with a TAR vendor, check references. This provides peace of mind the TAR methodology will be deployed in a legally defensible way.

- **Documentation**: The importance of documentation cannot be overstated. Counsel or the vendor (or both) must have a system for documenting every decision made during the life cycle of the review project — from seed set collection through review team training and privilege log generation.

- **Foreign Language Expertise:** Always be sure your technology-assisted review vendor is able to review foreign language sets accurately. Paying attorneys or temporary reviewers to review documents in foreign languages can be costly, so applying TAR to a multilanguage document set can produce even more cost savings.

BEST PRACTICES FOR EMPLOYING TAR

Garbage In, Garbage Out: It's a cliché for a reason — especially when working with TAR. Pay close attention to what is fed into the system, and make sure it's consistent. TAR applies the attorney's instructions to the entire data set, so it's important to make sure the data put into the system are as clean as possible.

- Ensure good, clean text input at the document level.

- In dealing with foreign language characters, make sure the text files that hold the document-level text are in UTF-8 format, so characters do not get converted into question marks or wingdings.

- Be aware of documents that have very little or no text. For example, an email from the CEO to the CFO that says "make it so" or "FYI" could be the most critical document in the entire case. But those phrases are not important independently. You can only establish the importance *in context* — in the attachments, the dates or preceding email messages. Be aware the less text in the document, the less text the algorithm has to work with and the less reliable your results could be.

- Some systems will assess not only text, but also the different metadata fields, such as the senders and recipients, time and date stamps, and even the document types. Counsel must understand and ask questions about these criteria to ensure TAR is applied appropriately in a given case.

- When you have "hiccups" in processing — for example, a document is thrown back because it's password-protected — the efficiency of TAR can decrease. Make sure processing exceptions have been addressed so you have competent text to input into the system.

- Differences in time zones, character sets or formats may cause a problem while implementing technology-assisted review. When these issues arise, attempts should be made to standardize the processing before training the system.

You Get What You *Inspect*, Not What You *Expect*: Once counsel is comfortable with the vendor, the process and the initial data set, he/she is ready to start the technology-assisted review process. It is important to trust the system at this point, but it is equally important to continue to verify the process is moving forward as expected. Quality control and algorithm adjustment are crucial for any technology-assisted review project. Be sure to:

- Sample all levels of responsiveness. Take a random sample of documents that the system has classified as being highly likely to be responsive and a sample of those it thinks are not likely to be responsive. When feasible, one should also sample documents the system could not definitely classify.

- Draw random samples across the corpus of data throughout the review project to ensure everything is staying on track.

- Run keyword searches to see what comes up as responsive. Cross-reference that against the TAR results.

- Perfect the algorithm. In an iterative fashion, work with the TAR vendor to adjust the system's training as appropriate.

Counsel should search, test and perform quality control until he/she is comfortable the system has classified the document set effectively. Err on the side of over-inclusion rather than under-inclusion. One can also use search terms in conjunction with TAR to ensure all responsive documents are found. TAR need not be used exclusively.

TAR TO THE RESCUE

It is no longer typical for attorneys to review manually an entire body of documents in a case from beginning to end. Technology-assisted review is here to stay, and attorneys who learn to leverage mathematical algorithms, sampling and keyword search are well-positioned for success. Attorneys must also keep up with case law, which is continuously changing and setting the parameters by which TAR can and should be used. TAR adds much-needed rigor
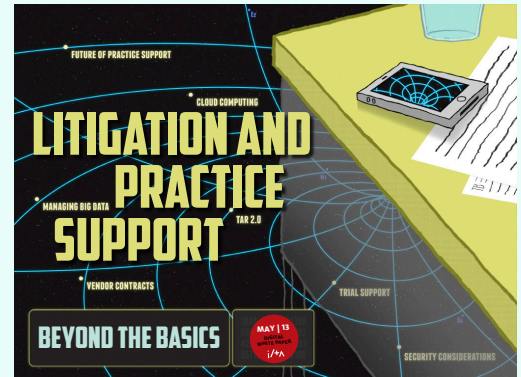
and efficiency to the review process, arming attorneys with the tools they need to deal with the exponentially growing volume of electronic documents. When deployed correctly and surrounded by reasonable, defensible processes, technology-assisted review can deliver speed, efficiency and accuracy to a traditionally cumbersome and expensive legal process. **ILTA**

Ignatius Grande is a Senior Discovery Attorney and the Director of Practice Support at Hughes Hubbard & Reed LLP. He can be contacted at grande@ hugheshubbard.com.

Andrew Paredes is the Senior Director of Document Review Services at Epiq Systems. He oversees the day-to-day operations of Epiq's document review services and assists law firms and corporate clients in the creation and implementation of review protocols, quality control and validation procedures. Contact him at aparedes@epiqsystems.com.