Bloomberg BNA

Digital Discovery & e-Evidence®

EMERGING TECHNOLOGY

Crowell & Moring's Jeane A. Thomas and Elizabeth A. Figueira explain why, despite initial skepticism about whether courts would accept technology assisted review and what costs and burdens would be required to defend its use, TAR now is, or should be, the mainstream approach for document review across a wide range of different types of matters.

Technology Assisted Review Goes Mainstream

BY JEANE A. THOMAS AND ELIZABETH A. FIGUEIRA

Technology assisted review (TAR) platforms, particularly those technologies commonly referred to as ''predictive coding,'' have been around for a number of years. Technology providers, along with early adopters, have promoted the potential for TAR to achieve significant efficiencies, cost savings and improved results over alternative review options including keyword searching and manual review.

Although there has been some concern about whether courts would accept TAR, and what costs and burdens would be required to defend its use, the benefits of TAR so dramatically outweigh those concerns that TAR now is, or should be, the mainstream approach for document review across a wide range of different types of matters.

Based in Washington, D.C., Jeane A. Thomas is co-chair of Crowell & Moring's E-Discovery & Information Management Group and is a partner in the firm's Antitrust Group.

Elizabeth A. Figueira is counsel in the New York office and is a member of the Litigation and E-Discovery & Information Management groups.

What is TAR?

In simplest terms, TAR involves the use of computer software to categorize documents consistent with human judgments.

A more formal definition of TAR from The *Grossman-Cormack Glossary of Technology-Assisted Review* (2013 Fed. Cts. L. Rev. 7) is:

> ''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. Some TAR methods use Machine Learning Algorithms to distinguish Relevant from Non-Relevant Documents, based on Training Examples Coded as Relevant or Non-Relevant by the Subject Matter Experts(s), 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.''

The Workflow. In a typical TAR workflow, one or a few human ''experts'' review and code a small sample of documents, which are then used to teach the computer how to code the larger universe of documents. There are a variety of alternative methods to teach the computer, but the common thread is that the computer learns from the human input and applies a consistent set of rules to code the larger data set. TAR provides two clear and obvious benefits:

 ISSN 1941-3882

(1) TAR is more consistent than human coding (accuracy) and

(2) TAR eliminates the need for human review of large volumes of electronically stored information (ESI) (efficiency and cost savings).

People + Process + Technology

TAR does not work its magic alone. The effective and defensible use of TAR requires human expertise, along with well-designed processes and workflows, in addition to the technology itself.

Human expertise is needed along two lines: (1) ''expert'' coders who teach the machine how to code, and (2) project managers who design the workflow and test, validate and track the results.

Since the quality of the machine coding is only as good as the human judgments on which it is based, TAR requires a different approach than traditional manual review. The most consistent results are achieved when there is only one, or a small number of expert coders, who collaborate on their determinations, so that the machine is not trying to resolve the inconsistent judgments of different human reviewers.

Furthermore, because the decisions of these expert coders are extrapolated across the entire document population, lawyers with the greatest knowledge of the subject matter and discovery requirements are the ideal candidates. Thus, senior lawyers on the case team are far more likely to be involved in the expert coding of the initial training and quality control sets than junior associates or contract attorneys.

More importantly, the results of TAR are dependent on many different judgments made along the way regarding workflow design. For example:

- How will the initial seed sets of documents be selected?
- How many training rounds of coding need to be conducted?
- What quality control processes will be utilized before the system is ''stabilized''?
- What measurements of recall (the percentage of responsive documents coded by the machine as responsive) and precision (the percentage of documents coded by the machine as responsive that are actually responsive) are acceptable?
- What additional non-TAR quality control methods can be used to improve those metrics?

Further, there are decisions that must be made after the machine coding has stabilized.

Will certain groups of documents be manually reviewed for responsiveness? Privilege? Key content?

The answers to these and many other questions will vary with the type of matter. It may be acceptable to forego manual review of certain categories of machine-coded ''responsive'' documents in a huge product liability matter, but that may be an unacceptable risk for documents produced in a criminal investigation. The critical point is that TAR requires expertise in project design and management every step of the way.

Accuracy, Cost Savings and Efficiency

Traditional manual review of documents has been treated as the ''gold standard'' for years despite little analysis of the quality of the results. Yet, the empirical evidence that exists suggests that human reviewers are far less than perfect, and teams of attorney reviewers, particularly large teams, will often make conflicting and inconsistent determinations of relevancy.

Keyword searching has gained in popularity as a way of sifting through large volumes of ESI to attempt to identify relevant material, and although results vary, studies show that keyword searching also can be far less than perfect in identifying relevant documents.

Recognizing the need to demonstrate the efficacy of TAR, there have been a number of studies comparing the results of TAR with those of manual review and keyword searching. These studies have consistently shown that TAR is more accurate than other review methods along the two main metrics—recall and precision. That alone should drive the adoption of TAR as a preferred approach to document review.

Senior lawyers are far more likely to be involved in the expert coding of the initial training and quality control sets than junior associates or contract attorneys.

First Hand Experience. In addition, TAR has proven to dramatically reduce costs and improve efficiency over traditional review. Based on more than a dozen matters in which Crowell & Moring has used TAR, our experience is that manual review—including expert coding, quality control, and analysis of documents the computer is unable to code—is required for typically less than 10 to 15 percent of the universe of documents collected, and sometimes far less.

For example, in a recent major document review, more than 2.5 Terabytes of ESI was collected, which was reduced to over 1 Terabyte (approximately 3.5 million documents) after de-duplication and date filtering.

Using a TAR workflow, less than 350,000 documents or less than 10 percent of the total collection after de-duplication and date filtering required manual review.

Based on a standard cost assumption of $1 per document, the result is roughly $3 million in savings over a traditional linear manual review process.

This example is consistent with other matters, in which typically 80 percent or more of a document population is machine-coded as non-responsive and does not need to be manually reviewed.

Additional Benefits. Reducing the pool of documents that must be manually reviewed results in not only cost savings, but also time savings. A project that could take months to manually review can now be done in weeks with the effective use of TAR.

  DDEE ISSN 1941-3882

TAR requires expertise in project design and management every step of the way.

Expanding Applications of TAR

TAR has proven to be effective across a broad range of matters, including discovery in civil litigation, government investigations and criminal matters. The technology has produced very good results and efficiencies with both small and large data sets, and with document populations containing sophisticated, highly-technical subject matter.

Although there will be times when linear manual review, keyword searching or other document review methods are appropriate, TAR should be at least considered as a viable option for many matters, either alone or in combination with other review methods.

Going Beyond Relevance Determinations. Further, TAR is not limited to making relevant/non-relevant determinations. The technology continues to evolve and can now be used for issue coding, identifying privileged material, and confidentiality classification. Moreover, TAR can be used in the context of early case assessment, or otherwise, to identify the most relevant or critical documents, including in the analysis of documents produced by other parties.

Forecast

Standard discovery orders will address confidentiality and privilege considerations, but parties eager to use TAR may benefit from more detailed stipulations that set forth the protocol to be used by each party.

Much of the judicial writing on TAR in recent months has been in the form of case management orders where the court permits the use of TAR and encourages various forms of cooperation various forms of cooperation between the parties. (See related story, infra.)

What the Orders Address. The most detailed orders discuss a wide range of topics, starting with the selection of a vendor, and how the parties will create the seed set of documents for training and what information they will exchange regarding those seed sets. The stated goal is to foreclose later disputes regarding the efficacy of the process.

The orders may also set out the threshold metrics for precision and recall and the methods by which the other party will be able to test the results.

For example, setting relevant metrics in advance allows the parties to represent at the end of discovery that more than ''x percent'' of the responsive documents were produced and that additional efforts are not likely to yield that elusive ''smoking gun'' that the opposing party may later claim was not produced.

Parties eager to use TAR may benefit from detailed stipulations that set forth the protocol to be used by each party.

Privilege Issues. Many of the same courts have also encouraged parties to seek greater protection against waiver of privilege with expansive orders that would allow a party to retain privilege claims even upon the intentional disclosure of privileged documents. The hope is that such non-waiver orders will allow parties to save costs and time by foregoing the traditional manual review for privileged documents, although parties may be reluctant to go down that path even with such an order.

While this form of cooperation and transparency may seem counter-intuitive for both clients and their counsel, it has been one of the key themes of the cases involving TAR. Judges want the parties to reach agreement and address these issues early on, before large costs have been incurred and before issues are presented to the courts in the form of disputes. Concerns about defensibility and unfairness diminish when there is a clear understanding among the parties and the court.