

# Towards a Synthesis of Judicial Perspectives on Technology-Assisted Review

---

By Julia L. Brickell and Peter J. Pizzi



*Julia L. Brickell is Executive Managing Director and General Counsel of H5, a company that blends specialized search expertise with technology to address discovery, litigation preparation, investigations, compliance, and corporate data governance. Ms. Brickell has many years of experience as a commercial litigator. Ms. Brickell is a member of the IADC Technology Committee, on the boards of Lawyers for Civil Justice and the American Institute of Information Technology Professionals, and on the faculty of Columbia University's Executive Master of Science in Technology Management. Peter J. Pizzi is the Managing Partner of Connell Foley's New York office. His practice focuses on business litigation, class action defense, internet and IP litigation, internal investigations, labor and employment law, cyber security and data privacy, and securities litigation. Mr. Pizzi has handled matters on behalf of corporate clients in a wide array of industries, including pharmaceuticals, finance services, computer technology, cosmetics, industrial equipment, and food service and food processors. He is a member of the IADC and of the Executive Committee of the New York State Bar Association's Commercial and Federal Litigation Section and Co-Chair of the Section's Internet and Litigation Committee. He formerly served as Co-Chair of the New Jersey State Bar Association's Internet and Computer Law Committee.*

---

**D**OCUMENT review accounts for an estimated 70% of all e-discovery costs. This means that document review also represents the greatest area of potential cost savings in e-discovery. Over the past ten years, the number of technological approaches to document review for litigation has increased, with the application of methods long in the domain of information retrieval. These advances call upon lawyers practicing civil litigation to gain a familiarity with the various technological op-

tions available, lest an opportunity is missed or the opponent, court, or client catches the unwary lawyer by surprise.

Technology-assisted review (“TAR”) is not a luxury available only to large firms handling very large cases. Properly chosen and deployed, technological methods can enable a smaller firm to handle larger cases, and thereby compete with larger firms. In the end, the goal is to make litigation more cost-effective, allowing more cases to stay in the judicial system rather than having litigants settle because the discovery costs

outweigh the value of the matter. In addition to gaining an understanding of the technology behind the available techniques, lawyers should appreciate that the match (or mismatch) between the problem to be solved, the technology chosen, and the expertise of the user are as important as the technology in determining whether the results will be both satisfactory and defensible.

Part I of this article explains TAR and the variety of approaches available. Part II synthesizes the meaning of the various judicial decisions dealing with TAR in a substantive fashion. Consideration of the judicial opinions proves the importance of understanding available technology. If there is an overall theme from the cases on this topic, it is that courts increasingly expect counsel to consider technological approaches and be competent to discuss what is warranted for a case. While the producing party still has the power to decide how to handle its production, if parties do reach agreement on an approach, they will be held to it, no matter how ill-advised it might turn out to be. Counsel must bring to the conversation sufficient expertise to understand the nuances and import of any proposed discovery protocol applying technology.

## I. Technology Assisted Review

Before 2000, review was done with large teams of document reviewers. Between 2000 and 2010, online review platforms entered the market, hosting whole cases on a single platform and reducing the transactional costs involved in having to manually pull or deliver documents. As court rules and courtroom lawyers began to focus on discovery of electronic information, research turned to

understanding how search could be applied to the review process. Research conducted from 2006-2011 under the auspices of the National Institute of Standards and Technology showed that TAR could significantly outperform or significantly underperform human review, depending on the tools used and the expertise of the users. An analysis of some of those results published in the *Richmond Journal of Law and Technology* (“*JOLT*”) concluded: “Overall, the myth that exhaustive manual review is the most effective – and therefore, the most defensible – approach to document review is strongly refuted. Technology-assisted review can (and does) yield more accurate results than exhaustive manual review, with much lower effort.”<sup>1</sup> The methods used by participants that returned top results ranged from sophisticated Boolean search to machine learning algorithms; some of the methods that performed less well fell into the same categories.

With the help of cases like *United States v. O’Keefe* and *Victor Stanley v. Creative Pipe* in 2008, the conversation about the process and expertise needed to successfully navigate document discovery in the increasingly electronic world of both individual and corporate clients entered the legal bar. Those foundational cases, followed by the *JOLT* article and the article “*Search, Forward*”<sup>2</sup> in *Law Tech-*

<sup>1</sup> Maura R. Grossman and Gordon V. Cormack, *Technology-Assisted Review In E-Discovery Can Be More Effective And More Efficient Than Exhaustive Manual Review*, 17 *RICH. J.L. & TECH.* 11, 48 (2011), available at <http://jolt.richmond.edu/v17i3/article11.pdf>.

<sup>2</sup> Andrew Peck, “*Search, Forward*” *Law Technology News*, October 2011 at 25, available at [https://law.duke.edu/sites/default/files/centers/judicialstudies/TAR\\_conference/Panel\\_1-Background\\_Paper.pdf](https://law.duke.edu/sites/default/files/centers/judicialstudies/TAR_conference/Panel_1-Background_Paper.pdf)

nology News authored by SDNY Magistrate Judge Andrew Peck, increased the awareness of the uses of search in this context. In parallel, amendments to ABA Model Rule 1.1 and increasingly, state rules of professional conduct, have focused on knowledge of technology (either hired or acquired) as an element of “competence” for lawyers.

With the safety of acquiescence (and at times encouragement) from the courts, the use of TAR has become more commonplace. TAR is sometimes called computer-assisted review (“CAR”) or content-based advanced analytics (“CBAA”) and is often referred to as predictive coding. It is important for lawyers to become familiar with the technology behind these techniques. The various forms of TAR in the market today have their origin in information retrieval techniques that have been the purview of doctoral theses and research for the past 50 years. In today’s market, two types of TAR predominate: approaches based on search terms and machine learning approaches.

Search based upon search terms is the easiest to understand: it is based on words you can read. It can range from simple keyword searches to complex search strings. User input comes in the form of building search terms that are tailored to the language the client uses in its discussion of the relevant subject matter; this language will likely differ by department and from client to client. Search terms can be tested and adjusted, with important words added and over-broad terms anchored or narrowed to home in on relevant content. The impact of these changes can be evaluated by running iterative searches on sample data, allowing the terms to be evaluated and refined. The effectiveness of the search will depend on the methodology

followed in designing and refining the search terms, the expertise and know-how of the individuals designing them, and the capability of the search tool to handle multiple or complex searches.

Machine learning approaches fall into several categories. What is common to the approaches is the use of an algorithm (a set of mathematical instructions) that builds a model of a class of documents based on various document features (chiefly words). Each technology vendor’s algorithm is different, counting and weighting the words and features in the documents in a particular way (perhaps ignoring some and weighting others more heavily, maybe or maybe not taking into account the order of the words). In practice, counsel neither sees nor adjusts the algorithm.

Machine learning approaches can be supervised or unsupervised. In a supervised approach, the user supplies the algorithm with a pre-labeled training set of documents on the basis of which the algorithm builds a model for classifying unseen documents. For example, a user might supply the algorithm with a sample of documents manually pre-coded as responsive and non-responsive; from that sample the algorithm would build a model for classifying documents not yet seen. The exact impact of the user input cannot necessarily be predicted or discerned, as it depends on the manner in which the input affects the model that the algorithm is building. The algorithm may rank the results based on a mathematical assessment of similarity to the algorithm’s model of the original input. Predictive coding, which has become a shorthand for any machine learning technique used for document review (but in fact is just a particular variation) is of this type. In an unsupervised approach, the algorithm is not

supplied with training data and instead tries to discover salient patterns in the document population. Many clustering tools (tools that attempt to categorize a document population by like content) are of this type.

The utility of different techniques depends on their being applied to solve a problem for which they are suited. Confronted with legacy media with unknown content, an unsupervised clustering algorithm could elucidate the prevalent topics in the data in short order. If the document requests focused on securities trading, for example, and the clustered topics in the legacy media related to company facilities, that media would not likely be a reasonable place to search for responsive content. In early case assessment, a machine learning technology, whether supervised or unsupervised, might enable a lawyer to delve into clusters of content likely to be pertinent, providing some order to the inquiry. In a document review context, a supervised learning approach could help counsel prioritize a manual review, and could be used by counsel to skip some segments of the data population entirely. In contrast, to pinpoint documents for production, an in/out decision could be made using a sophisticated approach based on search terms. The search terms can be designed by learning the language of the client from samples containing both relevant and nonrelevant documents. The effectiveness of the searches can be evaluated based on sample runs, and the search terms then iteratively refined and expanded based on results. The search terms can then be run by the software across the entire data population, tagging each document that contains the terms that make the document potentially relevant. These search results can provide a honed set for

counsel review or for quality assessment before production. The reduced volume in this targeted review set translates into reduced review hours, which, depending on the licensing costs or fees for services and time spent on user input, translates into cost savings for the client and frees up the budget for more important legal work.

Information science is playing a bigger role in search because it helps us evaluate the effectiveness of any search. The key metrics that measure effectiveness are “recall” (the percentage of responsive documents identified) and “precision” (the percentage of tagged documents that are actually responsive). Results from manual review, search terms, predictive coding, and/or any other search methodology can be evaluated by looking at these metrics.

In an adversarial proceeding where document review is required, TAR may be a topic of conversation at a meet-and-confer between counsel, pursuant to provisions like Federal Rules of Civil Procedure Rule 26(f). It is common for counsel to discuss keywords to be used to cull-in data, but counsel differ on their willingness to share their review plans. What about counsel’s duty to certify the accuracy of discovery responses?<sup>3</sup> While the proportionality principle set forth in Rule 26(b)(2)(C) refers to a limit set by a court, its cross reference in Rule 26(b)(1) – and its inclusion in the front of 26(b)(1) in the proposed amendments to the Rules – brings it into the ambit of document responses. Relatedly, the courts are increasingly interpreting the Rules for pro-

---

<sup>3</sup> See Rule 26(g)(1)(A) (using the term “complete and correct” with respect to a “disclosure” covered by Rule 26(a)) and Rule 26(g)(1)(B) (using the phrase “consistent with these rules and warranted by existing law [or nonfrivolous extension]” with respect to discovery responses).

duction to require reason, not perfection, in execution. Judge Peck, the first jurist to address predictive coding, opined that the proportionality principle requires a reasonable search, not “perfection.”<sup>4</sup>

While Rule 26(g)(1)(B) may not require perfection, the increasing focus of rules of professional conduct on the need for lawyers to hire or acquire technological competence suggests that poor execution due to ignorance is less and less likely to be tolerated. Using a tool for the wrong purpose or using it poorly is unlikely to be deemed a proportional response. Reaching agreement with opposing counsel, while providing protection from the opposition, is not likely to offer protection from a client who has been ill-served by an agreement made in ignorance. Sampling can be done well or poorly, statistics can be meaningful or meaningless. Properly executed, search results can be measured and substandard results improved. Rule 26(g) certifications, and proportionality, should equate to reasoned choices, properly implemented, to meet the actual needs of a case.

## II. Using Technology Assisted Review

As of this writing, there are approximately two dozen cases that address or bear upon the use of TAR methodologies. These cases often use predictive coding generically to refer to TAR and provide general guidance for all forms of TAR. A few cases touch upon these technologies in the course of approving an award of legal fees, or in showing the thoroughness of

counsel’s investigation into the events in question. Decisions that have delved into TAR in a substantive way will receive further scrutiny here.

There are common themes that emerge from what, at first glance, may seem to be discordant outcomes. All things equal, the producing party’s viewpoint regarding the search methodology to apply in order to retrieve responsive documents from its own collection generally prevails. As with all discovery issues, courts require parties to confer and are eager for lawyer cooperation to lead to agreement rather than have a search or review issue presented to the court for determination. If such an amicable resolution is reached, courts will usually hold both sides to the terms of their agreement. Not surprisingly, courts have limited capacity to get “into the weeds” and actually determine by judicial fiat the specifics of the TAR protocol to be employed in any particular case. Lastly, lawyers who decline to become versed in the available TAR options do so at their own peril and that of their clients. Courts may raise the use of TAR protocols *sua sponte* and increasingly expect lawyers to possess sufficient facility with technology to speak in an educated fashion not only about sources of electronically stored information but also about the utility of a TAR protocol for the matter at hand. If not sufficiently versed, lawyers will be ill-prepared to discuss or influence the components of a TAR protocol and risk agreeing to one of ineffective design.

*Da Silva Moore v. Publicis Group* is the first case to address whether predictive coding, as a form of TAR, comports with obligations imposed by the Federal Rules of Civil Procedure and whether the

---

<sup>4</sup> *Da Silva Moore v. Publicis Groupe SA*, 287 F.R.D. 182; 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. February 24, 2012 & April 25, 2012).

technique was suited to the particular controversy before the court. A proposed class of female executives sued Publicis Groupe, one of the “big four” advertising agencies, alleging sex discrimination, a “glass ceiling” and “systemic, company-wide gender discrimination.” (The case is still pending.) The issue was how defendants were to retrieve responsive documents from an anticipated set of three million. Magistrate Judge Peck, who had recently written his article *Search, Forward*, was referred the case. This assignment proved fortuitous for the defendants, who wanted to use predictive coding to ease the ESI cost burden. The plaintiffs’ putative class agreed to predictive coding and to a training protocol that included sharing of defendant’s training, or “seed,” sets. But in the end, agreement about a specific validation protocol eluded the parties. For example, plaintiffs challenged the validation protocol as lacking the detail needed to demonstrate that there would be meaningful measures of recall; they asserted that what was detailed – for example sample sizes for validation – would be quite inadequate to validate that the defendant achieved a high level of recall. Judge Peck ultimately accepted the protocol to which both parties had initially agreed, which continued to be supported by the producing party. This “final ESI protocol” (with footnoted objections) was filed on the docket and so-ordered by the court. Judge Peck allowed that after completion of the review, when results were available, if plaintiffs objected to the results, the reliability could be evaluated then.

On appeal to District Judge Carter, the court affirmed, noting that the plaintiff class had agreed upon the use of predictive

coding and the protocol itself had ample provisions permitting plaintiffs input and the magistrate judge had invited the plaintiff class to revisit the protocol should later document productions suggest problems.<sup>5</sup>

In *Da Silva Moore*, the parties reached agreement to use predictive coding, but when issues arose, the court held the parties to that agreement and incorporated the TAR protocol supported by the producing party, illustrating the first two principles that flow from TAR cases. *Da Silva Moore* demonstrates both judicial support for the use of TAR methods and the extensive detail regarding TAR with which the lawyers (with assistance of experts) need to be conversant. The case does not demonstrate the “right” sampling or training protocol, not only because the protocol was left open for later objection, but also because software, methodology, and validation must be appropriately matched to the needs of the situation being addressed.

A few months after *Da Silva Moore*, *Kleen Products LLC v. Packaging Corp. of America*<sup>6</sup> addressed predictive coding in the context of a price-fixing antitrust class action. The class plaintiffs challenged past document productions by defendants as inadequate and sought an order requiring the defendants to replace their Boolean search term methodology with a different form of TAR for increased accuracy. The defendants countered that the search strategy used for initial responses was honed by input from both sides, had been

<sup>5</sup> *Da Silva Moore v. Publicis Grp. SA*, 2013 U.S. Dist. LEXIS 155737 (S.D.N.Y. Oct. 30, 2013).

<sup>6</sup> Case: 1:10-cv-05711 (N.D.Ill. August 21, 2012)



subjected to QC testing, and had already yielded three million documents. The court signaled that it was not inclined to compel the defendants to redo work already performed with the participation of the plaintiffs. After urging from the Court, the parties resolved the issue by a stipulation which called for the application of TAR to future productions. *Kleen Products* illustrates again that the producing party's approach to its document production is typically permitted and counsel must have the technological competence to understand the consequences of discovery measures to which they agree because such agreements are not easily undone.

In the fall of 2012, a Delaware Chancery jurist, in *EORHB, Inc. v. HOA Holdings LLC*,<sup>7</sup> *sua sponte* directed the parties to show cause why TAR should not be utilized in a dispute involving a claim for indemnification arising out of the sale of the Hooters restaurant chain. Not only were the parties to utilize TAR, the Vice Chancellor directed them to utilize the same vendor, or to show cause why a joint selection was not feasible. Viewed on its own, the decision in *EORHB* would not fit easily into the rubric discussed above, but there is a post-script. Seven months later, the court revisited its predictive coding ruling at the request of the parties. The Vice Chancellor acceded to the parties' joint request to rescind that part of the order that required predictive coding because the volume of documents simply did not make infeasible the traditional document-by-document method of review.<sup>8</sup>

Similarly, the judge in the much-storied litigation arising out of the \$18 billion judgment obtained against Chevron in Ecuador by alleged victims of pollution caused by defendant's oil exploration suggested the use of predictive coding to reduce discovery costs. In *Chevron Corp. v. Donziger*,<sup>9</sup> Chevron alleged that the judgment was procured by fraud on the part of plaintiffs' lead counsel. Once this collateral litigation was in progress, Chevron served a non-party subpoena upon Patton Boggs, the law firm assisting the solo practitioner in representing the plaintiffs. (A year later, Chevron amended the complaint to add Patton Boggs as a direct defendant, and still months later, Chevron and Patton Boggs settled, clearing the way for the law firm merger forming Squire Patton Boggs.) Patton Boggs objected to the broad scope of the Chevron subpoena on numerous grounds, including, most notably, undue burden. Judge Kaplan was not impressed. Among other points made in his 73-page decision, Judge Kaplan pointed out that he had suggested Patton Boggs implement predictive coding to reduce the volume of documents it planned to review manually, but the firm gave the suggestion merely footnote attention, noting that the topic was under investigation:

At the September 2012 hearing, the Court urged the parties to analyze, in their subsequent submissions with respect to burden, whether and to what extent predictive coding could "reduce

<sup>7</sup> Case No. 7409-VCL (Del. Ch. October 15, 2012).

<sup>8</sup> *EORHB, Inc. v. HOA Holdings, LLC*, No. 7409-VCL, 2013 WL 1960621 (Del. Ch. May 6, 2013).

<sup>9</sup> Case 1:11-cv-00691-LAK-JCF (March 15, 2013).

the burden and effort” required to comply with the Subpoena. Apart from one footnote, PB’s submission ignored the subject entirely. The logical inference is that PB failed to address the subject because it would not have aided its argument.<sup>10</sup>

While the court seemingly suggested use of technology *sua sponte*, the ultimate decision on how to review the documents was left to the producing party. The court’s comment, however, may serve as a caution to lawyers: become familiar with technology where the magnitude of documents calls for its consideration; silence, for whatever reason, may not sit well with the court.

In *Hinterberger v. Catholic Health System*,<sup>11</sup> a Fair Labor Standards Act collective action, the parties for over one year had been unable to arrive at an ESI protocol covering the production of defendants’ documents using conventional keyword search methodology. The court urged the parties to consider predictive coding, referring them to Judge Peck’s decision in *Da Silva Moore*. In subsequent submissions to the court, the defendants advised the plaintiffs that they intended to use predictive coding and ultimately provided a protocol. Plaintiffs protested (by motion) that the protocol did not give plaintiffs access to seed sets used to train the software. The parties implicitly agreed that *Da Silva Moore* did not require a sharing of seed sets (as the *Da Silva Moore* parties had agreed to such sharing and the court merely confirmed that agreement),

but plaintiffs sought such sharing in the spirit of cooperation and forewarned of an ultimate challenge after production was completed. Defendants, the producing parties, maintained control of the process, confirming that they were aware of their discovery obligations, promising to meet and confer, and obtaining a dismissal of plaintiffs’ motion that objected to defendants’ protocol.

*Progressive Casualty Insurance v. Delaney*<sup>12</sup> follows the second theme discussed above, holding the producing party to its agreed protocol. The litigation was an insurance coverage action brought by an insurer against the FDIC as receiver. At stake were certain directors and officers policies procured by a group of failed banks. The parties arrived at an agreed protocol, approved by the court, to collect potentially responsive documents, assembling 1.8 million records. Pursuant to the protocol, the parties arrived at search terms, which reduced the number of documents to 565,000. The plaintiff insurer then had the option, under the protocol, to manually review the results or simply produce them, and was obligated to notify the FDIC of its choice. The plaintiff insurer began manual review, but then abandoned it. Despite the agreed protocol and representations that production was imminent, the plaintiff insurer then began to use predictive coding to further cull the review set. The plaintiff insurer did not first obtain approval from (or even provide notice to) the defendant FDIC before applying predictive coding to the results of conventional search, thereby unilaterally departing from the discovery protocol.

<sup>10</sup> *Id.* at 70 (slip op.).

<sup>11</sup> No. 08-CV-380S(F), 2013 BL 136456 (W.D.N.Y. May 21, 2013).

<sup>12</sup> Case No. 2:11-cv-00678-LRH-PAL, (Mot. Compel (Dkt. # 89)) 2014 BL 140634 (D. Nev. May 20, 2014).



The change was disclosed only after months of delay and in preparation for a status conference. The FDIC objected on several grounds, including to the insurer's proposed application of predictive coding on top of a traditional culling of documents through the application of search terms. The FDIC also argued that the insurance company failed to provide sufficient transparency during the history of interactions between the parties.

After a fruitless attempt by the parties to negotiate an amended protocol, the court ruled in the FDIC's favor, holding the plaintiff insurer to the original agreed protocol. Further, in view of discovery delays, perceived deficiencies in the new proposal to use predictive coding only on the results of search terms, the history of hotly contested discovery disputes, and the producing party's failure to advise the FDIC if it would review or simply produce the documents hit by the search terms, the court ordered the insurer to produce the entire 565,000 document set yielded by the application of search terms, subject to certain measures to protect privilege.

The magistrate judge in *Bridgestone Americas, Inc. v IBM Corp.*,<sup>13</sup> in contrast, allowed a Bridgestone to substitute a TAR technology for the agreed "attorney" review of documents hit by agreed keywords. The matter involved a dispute over the efficacy of an IBM software implementation and potentially involved scores of custodians and a timeframe of many years. The first 20 custodians had millions of emails that still numbered above 2 million after application of the agreed keywords. A few months earlier, during an extended keyword negotiation, the court had asked if

the parties had considered predictive coding. When Bridgestone disclosed its intent to use attorney-trained TAR to determine relevant documents, IBM argued that it was too late to switch. IBM noted that it was well into its own manual review for its production and objected that the technical know-how required, negotiation of protocols, and training of software would not save time or money. The court, mindful of the goal of efficient and cost-effective discovery and the volumes to be reviewed, permitted the change. The court noted that IBM is a "sophisticated user of advanced methods of integrating and reviewing large amounts of data" and made clear that IBM was welcome to switch to a TAR approach as well."<sup>14</sup>

In *re Biomet M2a Magnum Hip (Implant Products Liability) Litigation* (MDL 2391),<sup>15</sup> again shows the power of the producing party to control the dialogue regarding the search of its own documents. The plaintiff class had advised the defendant *not* to embark on document retrieval efforts until the case had gone through the MDL process and an agreed protocol was reached. The defendant proceeded nonetheless. It identified a data set of 19.5 million documents, used Boolean search terms to cull the collected documents (reducing the in-scope set to 3.9 million), then deduplicated the result to arrive at a review set of 2.5 million documents. To this reduced population it applied predictive coding and manual review and began producing the resulting documents. The plaintiffs believed that the production should have been in the range of 10

<sup>13</sup> No. 3:13-cv-01196 (M.D. Tenn. July 22, 2014).

<sup>14</sup> *Id.* at 2 (slip op.).

<sup>15</sup> No. 3:12-MD-2391, 2013 WL 6405156 (N.D. Ind. Aug. 21, 2013).

million documents. The plaintiffs insisted that the defendant apply predictive coding to the 19.5 million document set, noting its warnings to defer production efforts until a protocol could be agreed. The court rejected plaintiff's position as imposing an undue burden upon the defendant, not convinced that predictive coding over the entire 19.5 million documents would result in a more complete and accurate production than the defendant's unilateral efforts. Thus, the producing party's approach ultimately prevailed.

Most noteworthy in *In re Biomet*, perhaps, is that the plaintiffs were not able to demonstrate to the judge that the production was deficient or that the burden of its requested remediation was light. The key ingredient in the first was that the percentage of relevant documents in the discard pile *needed to be multiplied by the size of the discard pile* to understand the magnitude of what had been missed. The court noted that sampling had shown that between 1.37 % and 2.47% of the full collection was responsive, and that between .55% and 1.33% of the discard pile remained responsive, concluding that a "comparatively modest number" of responsive documents would be uncovered by running the predictive coding software over the discard pile.<sup>16</sup> In fact, however, it could well be that less than half of the responsive documents were produced. The sample showed between 267,150 and 481,650 responsive documents in the full collection and between 85,800 and 207,480 responsive documents in the discard pile. Using midpoints of the two ranges, it is estimated that no more than 60.8% of the responsive documents made it into the set to which predictive coding

was applied; some 40% of the responsive documents remained behind. And assuming that the predictive coding applied to the culled-in set achieved 70% recall (not an un-generous assumption), the recall of the production was no better than 42.6%, meaning that more than half the responsive documents were not produced. Regarding the burden, to rerun a pre-trained software over the discard pile would be a relatively light burden if the intent was not to re-train.

*Dynamo Holdings Ltd. v. Commissioner of Internal Revenue*,<sup>17</sup> is another case in this sequence, noteworthy because the party objecting to TAR was the United States Government, in the person of the Commissioner of Internal Revenue. At issue was whether certain transfers from one entity to a related entity were *bona fide* transfers for value or disguised gifts to the owners of the recipient entity. The Commissioner sought production of relevant information from within large quantities of backed-up email and file shares that also contained protected health information and other confidential information. In the alternative, the Commission sought delivery of the backup storage tapes on which the ESI was stored. The taxpayer objected that the request was a fishing expedition and that it would cost \$450,000 to review all of the documents on the tapes and determine which were responsive and which were not, leading the taxpayer to propose a predictive coding solution. After considering the testimony of the experts and reviewing the cases already passing on the use of TAR in litigation, the court rejected the Commissioner's position that TAR was an "un-

<sup>16</sup> *Id.* at \*5.

<sup>17</sup> 143 T.C. 9, 2014 WL 4636526 (T.C. Sept. 17, 2014).

proven technology.”<sup>18</sup> To the contrary, the court stated: “In fact, we understand that the technology industry now considers predictive coding to be widely accepted for limiting e-discovery to relevant documents and effecting discovery of ESI without an undue burden.”<sup>19</sup> The court ruled that the taxpayer’s choice of predictive coding was acceptable, pointing out, first, the anomaly of the court being asked to determine how the parties were to conduct the search for responsive documents:

And although it is a proper role of the Court to supervise the discovery process and intervene when it is abused by the parties, 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, such as whether that review should be done by a paralegal, a junior attorney, or a senior attorney. Yet that is, in essence, what the parties are asking the Court to consider—whether document review should be done by humans or with the assistance of computers.<sup>20</sup>

The court noted that if respondent ultimately finds the discovery response inadequate, respondent could file a motion to compel a better production.

Most recently, Magistrate Judge Peck returned to the subject of TAR in *Rio*

*Tinto Plc v. Vale S.A.*<sup>21</sup> The complaint in that case accuses the defendants of stealing mining rights in the Simandou region of southeast Guinea worth billions of dollars. Despite the exotic locale and allegations of racketeering and other wrongs, the parties reached an agreement to use TAR and submitted a TAR protocol to the court for signature. Judge Peck issued an opinion, citing the e-discovery community’s interest in the topic of TAR. Judge Peck concluded that in “the three years since *Da Silva Moore*, the case law has developed to the point that it is now black letter law that where the producing party wants to utilize TAR for document review, courts will permit it.”

An issue that remains unresolved is whether the “seed sets” – collections of documents used to train the software – are discoverable. By definition, the seed sets contain irrelevant as well as privileged documents, yet the seed sets are utilized as input to certain search algorithms that will result in the production of documents to the adverse party. In *Da Silva Moore*, Judge Peck (with the parties’ acquiescence) contemplated repeated meet-and-confer sessions regarding documents in the seed sets. In *Hinteberger v. Catholic Health System*, the court also appeared to contemplate interaction between the parties regarding seed sets. In *Biomet*, by contrast, the court rejected the plaintiff’s request for access to seed sets in rather blunt terms: “That request [for documents in the seed sets which were designated non-relevant] reaches well beyond the scope of any permissible discovery by seeking irrelevant or privileged documents used to tell the

<sup>18</sup> *Id.* at 3 (slip op.).

<sup>19</sup> *Id.* at 15 (slip op.).

<sup>20</sup> *Id.* at \*3-4.

<sup>21</sup> 1:14-cv-3042 (S.D.N.Y. Mar. 2, 2015). In the interest of disclosure, Judge Peck reviewed an earlier draft of this article.

algorithm what not to find. That the Steering Committee has no right to discover irrelevant or privileged documents seems self-evident.”<sup>22</sup>

### III. Conclusion

The jurisprudence on technology-assisted review shows that the producer most often decides the search method; agreements between parties control the discovery protocol and rarely will one party be

permitted to deviate unilaterally; and judges, while participating in extensive discussions and hearings and permitting voluminous briefing, prefer not to get involved in detailed decision-making that would supplant the view of the producing party. If lawyers come to court unprepared to explain available technologies and the results that may be expected from the various options, their clients will surely bear the consequences.

---

<sup>22</sup> 2013 WL 6405156 at \*2.