



5-23-2013

The Collision of the Courts and Predictive Coding: Defining Best Practices and Guidelines in Predictive Coding for Electronic Discovery

Elle Byram

Follow this and additional works at: <http://digitalcommons.law.scu.edu/chtlj>

Recommended Citation

Elle Byram, *The Collision of the Courts and Predictive Coding: Defining Best Practices and Guidelines in Predictive Coding for Electronic Discovery*, 29 SANTA CLARA HIGH TECH. L.J. 675 (2012).

Available at: <http://digitalcommons.law.scu.edu/chtlj/vol29/iss4/4>

This Article is brought to you for free and open access by the Journals at Santa Clara Law Digital Commons. It has been accepted for inclusion in Santa Clara High Technology Law Journal by an authorized administrator of Santa Clara Law Digital Commons. For more information, please contact sculawlibrarian@gmail.com.

THE COLLISION OF THE COURTS AND PREDICTIVE CODING: DEFINING BEST PRACTICES AND GUIDELINES IN PREDICTIVE CODING FOR ELECTRONIC DISCOVERY

Elle Byram[†]

TABLE OF CONTENTS

INTRODUCTION	676
I. THE EVOLUTION OF THE CURRENT SEARCH AND REVIEW STRATEGIES	677
II. TECHNOLOGY ASSISTED REVIEW: WHAT IT IS AND WHY IT HAS NOT YET BEEN UNIVERSALLY ADOPTED.....	682
III. TECHNOLOGY ASSISTED REVIEW'S CHANGING LANDSCAPE: RECENT COURT DECISIONS	685
A. <i>Moore v. Publicis Groupe</i>	685
B. <i>Global Aerospace v. Landow Aviation</i>	688
C. <i>Kleen Products v. Packaging Corporation of America</i> ..	690
D. <i>In re Actos (Pioglitazone) Products Liability Litigation</i>	692
E. <i>EORHB, Inc. v. HOA Holdings, LLC</i>	693
IV. REMOVING THE FEAR FROM TECHNOLOGY ASSISTED REVIEW: CLARIFYING BEST PRACTICES AND GUIDELINES.....	693
CONCLUSION	701

[†] J.D., UC Davis School of Law (2005); M.S. in Psychology, Pacific Graduate School of Psychology (now Palo Alto University) (2001); B.A. in Psychology, University of Colorado, Boulder (1996). Elle Byram is a project manager at SFL Data, where she sees electronic discovery from its inception through production. In her role as a project manager, Elle has worked with predictive coding tools to improve efficiencies and reduce costs in preparing large and small productions. Prior to her role at SFL Data, Elle was corporate counsel at ZL Technologies, a late-stage enterprise software company that develops archiving, e-discovery, records management and compliance software. Elle has also worked as a litigation attorney, managed and participated in numerous document review projects and recruited for agencies that place contract attorneys on e-discovery document reviews. She can be contacted at elle.byram@gmail.com.

INTRODUCTION

Technology assisted review (TAR), also known as predictive coding or computer-assisted review, has been a subject of serious debate in electronic discovery (“e-discovery”) for the past several years. For years, proponents of TAR have refrained from using the technology and instead were waiting for a judicial decision that would permit its use.¹ Fortunately, that decision has now arrived. In the first few months of 2012 the use of TAR has been directly addressed in three cases: *Moore v. Publicis Groupe*² and *Global Aerospace v. Landow Aviation*³ permitted TAR’s use. In *Kleen Products v. Packaging Corporation of America*,⁴ the parties and the court issued a stipulation and order in August of 2012 that set forth the parameters for electronic discovery, but did not mandate the use of predictive coding by the defendant.⁵ While proponents of TAR view *Moore* and *Global Aerospace* as wins for the use of TAR, the debate surrounding when and how it should be used is far from over. In July 2012, *In re Actos*⁶ also reaffirmed the use of TAR. Most recently, in October 2012, Vice Chancellor J. Travis Laster in Delaware Chancery Court required the parties in *EORHB, Inc. v. HOA Holdings, LLC* to use TAR from the same vendor.⁷

TAR is an important next step in the development of e-discovery. The costs and inefficiencies resulting from the massive volumes of electronic data churned out by corporations on a daily basis have prohibited many businesses from truly litigating the substance of their disputes. Because discovery—and more importantly e-discovery—is a necessity of litigation in the United States, legal practitioners will need to rely on advances in technology to make discovery more manageable. TAR is one of these advances.

1. Andrew Peck, *Search, Forward: Will Manual Document Review and Keyword Searches be Replaced by Computer-Assisted Coding?*, LAW TECH. NEWS, Oct. 1, 2011, at 25, 29, available at http://www.recommind.com/sites/default/files/LTN_Search_Forward_Peck_Recommind.pdf.

2. *Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012).

3. *Global Aerospace Inc. v. Landow Aviation, LP*, No. CL 61040 (Va. Cir. Ct. Apr. 23, 2012) (Order Approving the Use of Predictive Coding for Discovery).

4. *Kleen Prods., LLC v. Packaging Corp. of Am.*, 775 F. Supp. 2d 1071 (N.D. Ill. 2011).

5. *See Stipulation and Order Relating to ESI Search*, *Kleen Prods. LLC v. Packaging Corp. of Am.*, No. 1:10-cv-05711 (N.D.Ill. Aug. 21, 2012).

6. *In re Actos (Pioglitazone) Prods. Liability Litig.*, No. 6:11-md-2299, 2012 WL 6061973 (W.D. La. July 27, 2012) (Case Management Order).

7. *EORHB, Inc. v. HOA Holdings LLC*, No. 7409-VCL, at 66 (Del. Ch. Oct. 15, 2012) (hearing transcript).

By using TAR, parties will be able to cull and review data more quickly and at a fraction of the cost of linear review.⁸ This promises to ease the burdens e-discovery places on the judicial process by reducing the number of discovery disputes on the court's dockets and thus allowing litigation to be resolved more efficiently and economically.⁹

This article explores the recent precedent and commentary surrounding use of TAR. While TAR is technology that should be embraced by the legal community, determining when its use will be beneficial and how it should be carried out still needs further clarification. Determining when it should be used requires an evaluation of all aspects of a case: the costs, the parties, the amount and types of data, the time for completion, and the type of review anticipated. Even when its use would be beneficial, the standards for how to use TAR must be set forth clearly. By clarifying the standards and best practices that should be used, parties will be better able to evaluate their cases to determine when and how to use TAR and thus be able to take advantage of the benefits it offers.

I. THE EVOLUTION OF THE CURRENT SEARCH AND REVIEW STRATEGIES

Searching for and reviewing documents has been a burdensome necessity of litigation for as long as modern day legal practice has been in existence. With corporations becoming larger and more complex, this burden has only increased. Data has become more varied and voluminous—beginning with paper and now largely electronically stored information (ESI).¹⁰ Before e-discovery, a typical large-scale document review required a team of junior associates and paralegals to spend weeks or months sitting in a dusty warehouse flipping page-by-page through boxes of documents. Records would be marked with flags and sticky notes to denote which were relevant. Privileged materials would be treated similarly and

8. A linear review consists of a record-by-record review of every record produced, usually performed by contract attorneys or legal professionals.

9. See, e.g., Edward M. McNally, *Delaware's Complex Civil Litigation Court: One Year Later*, DEL. BUS. LITIG. REPORT (May 18, 2011), <http://www.delawarebusinesslitigation.com/2011/05/articles/case-summaries/delawares-complex-civil-litigation-court-one-year-later/> (noting that electronically stored information resulted in escalated costs and delays in resolving disputes).

10. See Douglas W. Oard et al., *Evaluation of Information Retrieval for E-Discovery*, 18 ARTIF. INTELLIGENCE & L. 347, 348-49 (2010).

were ultimately redacted using white redaction tape. Reviewers' hands would end up cracked and dried. All of this occurred with hands and eyes—awake or not—touching each and every record.¹¹

As technology has evolved and paper records shifted to ESI, the task of locating and reviewing relevant data has become more cumbersome, time-consuming, and costly. Although ESI has removed the dusty warehouse from the equation, the volume and complexity of the data has increased the time and effort required to locate and review it.¹² Additionally, searching for and reviewing ESI has required the development of new strategies and software. Numerous vendors developed a variety of search tools that flooded the e-discovery market in its first decade—from about 2000 until 2010. The goal was to collect the universe of ESI and run searches on it, typically using keywords, to narrow the data to what was hopefully the most relevant material.

Keyword searching was, and still is, the primary method for searching and narrowing ESI to identify that which must be reviewed.¹³ Most keyword searching is done using Boolean operators, such as “and,” “or,” and “not,” as well as proximity operators (e.g. “w/3” meaning “within three words” or “w/s” meaning “within same sentence”)¹⁴ and wildcards or stemming devices.¹⁵ Many legal professionals are familiar with these types of searches from conducting legal research on Lexis-Nexis and Westlaw. However, as many associates know, the task of determining the precise words to find the perfect case, or for that matter to merely reduce the cases to a manageable amount for review, is difficult.¹⁶ Finding the right

11. It is rare to see a large-volume paper review of records. Records that are stored as only paper records are generally scanned as images and loaded to an electronic review platform.

12. See generally *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 8 SEDONA CONF. J. 189 (2007) [hereinafter *Sedona Conference*].

13. See George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J.L. & TECH. no. 3, at 1, 21 (2007).

14. A Boolean search is a method of searching using operators such as “AND” or “W/2” to help define relationships between the search words or other parameters. For example, AND will find two terms such as “Wine AND Vineyard”. “W/2” represents the distance between two words in the search and will find wine within two words of vineyard when “Wine W/2 Vineyard” is used. See, e.g., *Definition of: Boolean*, MERRIAM-WEBSTER, <http://www.merriam-webster.com/dictionary/boolean> (last visited Apr. 20, 2013).

15. Paul & Baron, *supra* note 13, at 22.

16. See Herbert L. Roitblat, Anne Kershaw & Patrick Oot, *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 J. AM. SOC'Y FOR INFO. SCI. & TECH. 70, 71 (2010).

keywords to use might not even be possible for an associate unless she is well-versed in the subject area.

These difficulties only increase when searching volumes of ESI in discovery. The colloquialisms and abbreviations different parties use in their documents make it many times more difficult to search these documents than a typical Lexis-Nexis search.¹⁷ Human language is inherently ambiguous, especially as the data is moved further away from the formalities that characterize data in more structured databases such as Lexis-Nexis.¹⁸ Indeed, the objective of Lexis-Nexis is to make it as easy as possible for the associate to locate that perfect case. Not so with corporate ESI. Most employees don't draft emails or prepare documents with the prospect of discovery guiding how the records should be drafted. Litigation is usually far from the author's mind.

In addition, the fast-pace of electronic communications has increased the creation of new language styles, words, and, ultimately, misspellings.¹⁹ One example that is rife with concerns is text messaging. Due to the shortened amount of space for a standard text message, words and phrases are frequently abbreviated. When drafting a quick email or using a chats system, words can easily be made up on the fly.²⁰ And, those quick back-and-forth emails are frequently not proof-read, resulting in numerous spelling errors. Dialects and uses of language will also vary in different locations and for different generations or ethnic groups.²¹ The variety of communication styles ensures that no two sets of ESI will ever conform to exactly the same standards.

As if the inherent problems of communication were not enough, the parties to litigation also need to reach agreement on numerous aspects of how to conduct electronic discovery. For example, what should the keywords be? A party who is uninformed about another party's records may have a difficult time determining what words the other party used.²² Associates would have a difficult time overcoming

17. See Paul & Baron, *supra* note 13, at 23; *Sedona Conference*, *supra* note 12, at 193, 202.

18. Paul & Baron, *supra* note 13, at 23.

19. *Sedona Conference*, *supra* note 12, at 202.

20. Paul & Baron, *supra* note 13, at 23-24.

21. *Id.* at 24.

22. David C. Blair & M.E. Maron, *An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System*, COMM. OF THE ACM, Mar. 1985, at 289, 293 (finding that recall was only 20% of the relevant documents).

this challenge without the other party first producing documents. The challenge continues to multiply as the number of parties and issues in a case increase. If the chosen keywords are too broad, the resulting dataset will be unwieldy and rife with non-relevant material.²³ Too narrow and the returned records may not include critical relevant information.

Technical factors can also affect keyword searching. Take, for example, optical character resolution (OCR).²⁴ OCR is a method in which software pulls text from records to create searchable text. This is very common with paper records where it is necessary to scan the paper document to turn it into a searchable electronic record. Some electronic records, such as non-searchable PDFs, must also have OCR applied because the text is not extractable. OCR becomes problematic for records that are not perfectly clear. Faxed and handwritten records frequently do not OCR well because the text is not perfectly clear and, therefore, OCR will not be able to duplicate the text resulting in documents that are difficult to search.

Despite the drawbacks inherent in a keyword search,²⁵ they still have value to the legal profession in some contexts. It can be an excellent way of searching smaller datasets and finding specific records, if an associate knows exactly what she is looking for. The more precisely the search terms are defined and the narrower the universe of data that is being searched, the more likely keyword searching will be successful.²⁶

In addition to keyword searching, alternative search methods, such as concept searching, have also been developed.²⁷ These methods use semantic or statistical relationships (or a hybrid of the two) between words to locate specific data. Concept searching, for example, would know that vineyard and grape are related to wine,

23. *Sedona Conference*, *supra* note 12, at 205.

24. Paul & Baron, *supra* note 13, at 24.

25. *See id.* at 24-25 & n.103.

26. *See, e.g.*, Sonya L. Sigler, *Are Lawyers Being Replaced by Artificial Intelligence? Moving Beyond Keyword Search: An Introduction to Advanced Search & Retrieval Technologies*, DESI III GLOBAL E-DISCOVERY/E-DISCLOSURE WORKSHOP AT THE 12TH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW (ICAIL 2009) (June 8, 2009), available at http://www.law.pitt.edu/EDSI#_Workshop/Papers/DESI_III.SSigler_final.pdf.

27. Jason R. Baron, *Law in the Age of Exabytes: Some Further Thoughts on 'Information Inflation' and Current Issues in E-Discovery Search*, 17 RICH. J.L. & TECH. no. 3, at 1, 25-26 (2011). A good resource outlining alternative search tools and methods is provided in *Sedona Conference*, *supra* note 12, at 202-03, 217-223.

even though the word wine might not be used.²⁸ Types of linguistic searching include taxonomies and ontologies and statistical search methods include clustering and latent semantic indexing. Alternative search methods tend to be used less frequently than keyword searching for many of the reasons that are discussed in the next section.²⁹

Searching is only a start. After the records have been searched, the data must still be reviewed. In e-discovery's early days, searching involved a linear approach—reviewing each and every document collected. This method of review is still viewed as the “gold standard” by many attorneys.³⁰ Similar to filing through boxes in dusty warehouses, teams of associates and paralegals sit in front of computers sometimes for fifteen hours a day or more reviewing documents. The reviewer applies coding decisions about relevancy, privilege, and even substantive issues. Through the industry's evolution, this has become the bane of contract attorneys, who in the early days were frequently billed by law firms at associate rates. Some of these reviews could last years and cost companies hundreds of thousands to millions of dollars.³¹

Time and costs are not the only factors that have made linear reviews problematic. There is oftentimes tension between those who are reviewing the data and those who are substantively involved with the case. Most large reviews and many small ones are frequently performed by swarms of contract attorneys. Their only guidance is from a junior associate who may be reluctant to lead the contract attorneys. Involvement by senior associates or partners is limited. The contract attorneys are frequently provided no more information than what is necessary—perhaps a brief summary of the case, the

28. *Sedona Conference*, *supra* note 12, at 202.

29. See Paul & Baron, *supra* note 13, at 26. “[T]here appears to be a myth that manual review by humans of large amounts of information is as accurate and complete as possible—perhaps even perfect—and constitutes the gold standard by which all searches should be measured.” *Sedona Conference*, *supra* note 12, at 199.

30. *Sedona Conference*, *supra* note 12, at 199. See also Patrick Oot, Anne Kershaw & Herbert L. Roitblat, *Mandating Reasonableness in a Reasonable Inquiry*, 87 DENV. U. L. REV. 533, 545 (2010).

31. As the industry has evolved, many companies have hired contract attorneys through staffing agencies, directly reducing the costs of reviews substantially. Despite these attempts to reduce review costs, review still comprises 73% of all e-discovery costs according to the RAND Institute study. See NICHOLAS M. PACE & LAURA ZAKARAS, WHERE THE MONEY GOES: UNDERSTANDING LITIGANT EXPENDITURES FOR PRODUCING ELECTRONIC DISCOVERY xiv-xv (2012), available at http://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1208.pdf.

complaint, and the production request.³² Additionally, long hours and tedious work can lead to reviewer fatigue.³³ Moreover, human nature can create inconsistency in coding decisions.³⁴ If the gold standard is to have eyes on each document, there seems to be a disconnect: Contract attorneys who lay eyes on each record are seldom involved in substantive decision-making in the case; yet, those involved in the substantive decisions are rarely the ones to view each record.

Because of the inherent difficulties with searching vast arrays of data that will continue to grow and morph for decades to come, many industry professionals have looked for alternative approaches, like TAR, for searching and reviewing what will soon be petabytes of ESI. Despite its early promise, numerous factors have prevented TAR from becoming the new gold standard in ESI.³⁵

II. TECHNOLOGY ASSISTED REVIEW: WHAT IT IS AND WHY IT HAS NOT YET BEEN UNIVERSALLY ADOPTED

Technology assisted review aids the search and retrieval process³⁶ by assisting with coding records with some designation, such as responsiveness or privilege.³⁷ TAR has also been referred to as predictive coding and computer-assisted review, among other terms. Some TAR programs, such as predictive systems—which will be the focus of the remainder of this article—apply coding decisions made by contract attorneys on a small subset of records to the remaining un-reviewed records.³⁸ Other systems, frequently called automated predictive coding, automatically tags the records based on user input but without reviewing and coding the records.

32. See, e.g., Roitblat et al., *supra* note 16, at 77.

33. *Id.*

34. Maura R. Grossman & Gordon V. Cormack, *Inconsistent Assessment of Responsiveness in E-Discovery: Difference of Opinion or Human Error?*, DESI IV: ICAIL 2011 WORKSHOP ON SETTING STANDARDS FOR SEARCHING ELECTRONICALLY STORED INFORMATION IN DISCOVERY PROCEEDINGS, at 9-10 (June 6, 2011), available at <http://www.umiacs.umd.edu/~oard/desi4/papers/grossman3.pdf>; Roitblat et al., *supra* note 16, at 77.

35. See, e.g., Sonya L. Sigler, *Permission Is One Thing; Adoption Quite Another*, THE NAT'L L.J., May 21, 2012.

36. The search and retrieval process is an early step in e-discovery process where data, after it is collected and processed, is searched to cull the data into manageable populations, but also to gain a more general understanding of the data as well.

37. Baron, *supra* note 27, at 26-27; Maura R. Grossman & 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. no. 3, at 1, 3 (2011).

38. Baron, *supra* note 27, at 26-27; Grossman & Cormack, *supra* note 37, at 3.

The concepts behind TAR are not new. Information retrieval has been burgeoning for over 50 years, but until recently it has been mostly of interest to those in information and library science.³⁹ Blair and Maron conducted the earliest study into information retrieval in the legal field in 1985.⁴⁰ This study reviewed the ability of search teams and reviewers to find relevant data. They found that manual search could not be relied on to retrieve records. Along with the recent Internet boom, more companies, including Internet search companies, have begun honing in on how to effectively search through hundreds of thousands (or more) of records.⁴¹ Improvements in search and information retrieval combined with increases in the production of ESI have made many legal practitioners question whether manual review is the best mode for reviewing and producing ESI.⁴²

TAR programs were first introduced to the e-discovery market during the e-discovery boom.⁴³ Most of the today's programs operate using algorithms derived from alternative search methods, such as conceptual searching or clustering, by looking for relationships between words in the data set. The collected data, either before or after it is culled or searched using keywords, is deposited into the software program.⁴⁴ From there, for the predictive models, a small team of associates and contract attorneys tag samples of those records for relevancy.⁴⁵ TAR programs then use the attorneys' decisions to categorize the remaining data that was not reviewed.⁴⁶ Samples are taken and reviewed by the attorneys. Any new decisions are fed to the program and used to further categorize the records. This iterative process is referred to as "training" the computer and continues until the team is satisfied that all of the records have been properly tagged.

Documents are typically categorized by their similarity. Some of the programs will give a percentage of similarity to the reviewed

39. *Sedona Conference*, *supra* note 12, at 204.

40. *See Blair & Maron*, *supra* note 22.

41. *See Oard et al.*, *supra* note 10, at 364.

42. *See, e.g.*, Gordon V. Cormack, Maura R. Grossman, Bruce Hedin & Douglas W. Oard, *Overview of the TREC 2010 Legal Track*, THE NINETEENTH TEXT RETRIEVAL CONFERENCE (TREC) PROCEEDINGS (2010), available at <http://trec.nist.gov/pubs/trec19/papers/LEGAL10.OVERVIEW.pdf>; Grossman & Cormack, *supra* note 37; Roitblat et al., *supra* note 16.

43. KERSHAW & HOWIE, *supra* note 27, at 6-10.

44. *See id.* at 6-10 ("Overall Process" column).

45. *See id.*; Grossman & Cormack, *supra* note 37, at 4.

46. *See KERSHAW & HOWIE*, *supra* note 27, at 6-10 ("Overall Process" column).

records, which is based on the instructions given to the program. Depending on the program used, the system can either provide a responsive or not-responsive determination, what is known as binary retrieval, or it can provide a ranking of how similar the records are to each other. What cut-off level a team uses to determine what to review or sample from those that are responsive will be based upon the available resources, the parties' decisions, and the needs of the case.⁴⁷

Recent research supports the benefits that are being sought by using TAR. TAR produces results that are, at a minimum, comparable to manual review.⁴⁸ The vast majority of disagreements in manual review occur due to human error rather than because documents fall in some grey area.⁴⁹ Human error can arise from an individual making inconsistent judgments or two or more separate professionals making decisions that are inconsistent with each other. TAR eliminates this variability by providing consistent judgments across the dataset.⁵⁰ Where the review was done properly, the research supports that TAR finds more responsive records.⁵¹ TAR also generally performs more efficiently and at a fraction of the cost than manual review.⁵²

Despite its benefits, attorneys have been reluctant to adopt TAR, holding firm to the belief that manual review is still the gold standard.⁵³ Despite research that supports inconsistent judgments by humans, some in the legal industry believe that information needs are more precisely defined for e-discovery than for classical information retrieval.⁵⁴ In essence, these legal practitioners believe that lawyers are better able to assess relevance and privilege than non-lawyers. They argue that the most defensible way to ensure a production is accurate is to have a lawyer review every document.⁵⁵ These beliefs contradict the findings: Independent assessors have difficulty

47. *See id.*

48. *See, e.g.,* Grossman & Cormack, *supra* note 37; Roitblat et al., *supra* note 16; Cormack et al., *supra* note 42.

49. Grossman & Cormack, *supra* note 34; Roitblat et al., *supra* note 16, at 77.

50. Grossman & Cormack, *supra* note 37, at 43; Roitblat et al., *supra* note 16, at 77-78.

51. Grossman & Cormack, *supra* note 37, at 43; Roitblat et al., *supra* note 16, at 79.

52. PACE & ZAKARAS, *supra* note 31, at 66-69.

53. *See generally* Grossman & Cormack, *supra* note 34.

54. *See* Blair & Maron, *supra* note 22, at 293 (finding that the lawyers searching for the records believed they were retrieving a much higher percentage of relevant records). *See also* William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 135 (S.D.N.Y. 2009).

55. *See* Grossman & Cormack, *supra* note 37, at 3. *See also supra* note 29.

applying relevancy consistently even when the information need is specified by attorneys.⁵⁶

Some resistance to TAR may arise from a lack of understanding the technology.⁵⁷ According to the Sedona Conference Best Practices on Search and Retrieval, many legal professionals hold the belief that computers cannot be programmed to replace the human intelligence needed to make complex determinations.⁵⁸ Some argue that a lack of scientific validity supporting TAR will preclude a successful defense of it in court.⁵⁹ Others believe that TAR will not be able to flag important documents, such as privileged documents or smoking guns. And still some fear that it will not result in a production of *all* the responsive records.⁶⁰ Regardless, the lack of understanding by attorneys of what technology can and cannot do will act as an impediment to the industry. Uninformed lawyers will make uninformed judgments and create unrealistic expectations by the courts.⁶¹

Some attorneys do embrace TAR, but despite being informed of TAR's potential benefits, these attorneys have been waiting for judicial guidance.⁶² They have declined proceeding with its use, choosing to avoid being the first lawyer to try it in court. This may be due to a fear of client reprisal or a bad result in court leading to sanctions. Considering that courts have lost patience for e-discovery guffaws, it is understandable that these attorneys would want to wait for someone else to adopt the technology, test it out in court, and receive the courts' affirmation.⁶³ The courts' recent decisions, however, provide this affirmation.

III. TECHNOLOGY ASSISTED REVIEW'S CHANGING LANDSCAPE: RECENT COURT DECISIONS

A. *Moore v. Publicis Groupe*

In February 2012, Judge Andrew Peck of the U.S. District Court

56. Grossman & Cormack, *supra* note 37, at 9.

57. *Sedona Conference*, *supra* note 12, at 203.

58. *Id.*; see also KERSHAW & HOWIE, *supra* note 27, at 30.

59. *Sedona Conference*, *supra* note 12, at 203; see generally Sigler, *supra* note 26.

60. PACE & ZAKARAS, *supra* note 31, at xviii.

61. Oot et al., *supra* note 30, at 545.

62. PACE & ZAKARAS, *supra* note 31, at xix, 77; KERSHAW & HOWIE, *supra* note 27, at 30.

63. PACE & ZAKARAS, *supra* note 31, at xix, 77.

for the Southern District of New York approved the use of TAR in *Moore v. Publicis Groupe*.⁶⁴ *Moore* is an employment class action dispute brought by five female plaintiffs against defendant Publicis Groupe.⁶⁵ Plaintiffs claimed that defendant had a “glass ceiling” that prevented women from advancing beyond entry-level positions and that there was system-wide gender discrimination. This is also the first lawsuit in which a court directly addressed the use of TAR, and it is the first case in which the court ordered its use.

In the dispute’s short life, the parties held numerous discovery conferences, which included discussions about the use of TAR. Due to the large volume of data, the parties agreed to its use.⁶⁶ Plaintiffs noted, however, that they had multiple concerns about how defendant MSL Group, a subsidiary of Publicis Groupe, was planning on using TAR, and wanted further clarification.⁶⁷ Notwithstanding, the parties moved forward with plans to use TAR to cull and ultimately reduce the amount of data that would be needed for review.

In anticipation of TAR’s use, defendant MSL Group provided a detailed protocol outlining each step that it would take.⁶⁸ The parties would draw and tag a random seed set of 2,399 records from the corpus of data.⁶⁹ In addition, to identify documents responsive to several of plaintiffs’ requests, MSL performed judgmental coding to further aid in training the tool. The parties conducted seven iterative rounds of sampling the data and training the tool. The protocol further stated that they only had to produce the top 40,000 records.

Notwithstanding the detailed protocol, plaintiffs were dissatisfied and filed objections. They argued that MSL’s protocol was not sufficiently transparent and lacked the necessary standards to be able to assess whether the results were accurate.⁷⁰ They also argued that it was arbitrary to allow only the top 40,000 records to be produced. What if substantially more were found to be responsive?⁷¹

64. *Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012).

65. *Id.* at 183.

66. With over three million emails to review, linear manual review was “simply too expensive.” *Id.* at 190.

67. *Id.* at 185.

68. *Id.* at 193-205 (Parties’ Proposed Protocol Relating to the Production of Electronically Stored Information (ESI) & Order).

69. The random sample of 2,399 records represented “a 95% confidence level with a confidence estimation of plus or minus 2%.” *Id.* at 200.

70. *Id.* at 189; Plaintiffs’ Rule 72(a) Objection to the Magistrate’s February 8, 2012 Discovery Rulings at 14-15, *Moore*, 287 F.R.D. 182 (No. 11-cv-1279).

71. *Moore*, 287 F.R.D. at 185.

Another objection was that plaintiffs could not determine whether defendant's method would work because there was no description of how many relevant documents could be allowed in the final random sample of irrelevant documents. In essence, because recall and precision—measures of completeness and accuracy respectively⁷²—were not sufficiently defined, the court was simply delaying answers to questions that could resolve a future argument.⁷³

Judge Peck ruled substantially in favor of MSL. Plaintiffs succeeded only in removing the limit on the number of records to be produced. The rest of their arguments were not successful. First, Judge Peck found that the defendant's method was transparent: Plaintiffs would be able to view every record in the seed set to evaluate whether it was tagged relevant or not relevant.⁷⁴ Any issues that needed to be resolved surrounding the tagging decisions could be brought to the court's attention. With regard to any relevant records found among those tagged "not relevant," the court held that a decision regarding the standards was premature; what action needs to be taken would depend on the results and the costs that would be incurred.

Judge Peck ended his February 22, 2012 order with several recommendations.⁷⁵ First, the time at which a review and production using TAR concludes will depend on when the software has been trained and the results have been verified. Second, to control discovery costs, the parties should begin with only the key custodians, even if this would result in a longer discovery process. Third, the more transparent a party is with his key custodians, and how he proposes to search for the records, the more likely the other parties and the court will be more willing to agree with the approach. Finally, hiring e-discovery vendors to be present at hearings or conferences and ready to provide their guidance will be beneficial. He concluded by noting that "computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases

72. See generally Grossman & Cormack, *supra* note 37.

73. Recall and precision are two measures of information retrieval. Recall is a percentage based on the number of documents retrieved divided by the number of responsive documents overall. Precision is the number of responsive documents retrieved divided by the number of documents retrieved. For a more thorough discussion on measuring information retrieval see *Sedona Conference*, *supra* note 12, at 205. See also Jason R. Baron, *Toward a Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery*, 6 SEDONA CONF. J. 237, 242 (2005).

74. See *Moore*, 287 F.R.D. at 189.

75. *Id.* at 192-93.

where it may save the producing party (or both parties) significant amounts of legal fees in document review.”⁷⁶

Judge Andrew Carter upheld Judge Peck’s order on April 26, noting that no review tool guarantees perfection.⁷⁷ Judge Carter stated: “[T]he ESI protocol contains standards for measuring the reliability of the process and the protocol builds in levels of participation by Plaintiffs. It provides that the search methods will be carefully crafted and tested for quality assurance, with Plaintiffs participating in their implementation.”⁷⁸ He continued: “At this stage, there is insufficient evidence to conclude that the use of the predictive coding software will deny Plaintiffs access to liberal discovery.”⁷⁹ MSL Group was granted the right to use TAR.

B. Global Aerospace v. Landow Aviation

As with *Moore*, TAR was also permitted in *Global Aerospace Inc. v. Landow Aviation, L.P.*⁸⁰ On April 23rd, 2012, in a one-page opinion, Circuit Judge Chamblin permitted its use in this data-sensitive matter. The matter arose from the collapse of three hangars at the Dulles Jet Center during a major snowstorm in February of 2010.⁸¹ Plaintiffs sued defendants claiming that design and construction deficiencies contributed to the hangar’s collapse. Shortly after filing its complaint, plaintiffs proceeded with discovery by filing requests for production, which included ESI. Defendant Landow took steps to collect and preserve all potentially relevant ESI, which resulted in more than eight terabytes of data and 250 gigabytes of reviewable ESI.

Due to the large size of data, Landow conducted a cursory preliminary review of the data to better understand the collection.⁸² Landow loaded some of the data into an e-discovery review tool to analyze the records, which separated the records into clusters. From their brief analysis, it became readily apparent that a majority of the

76. *Id.* at 193.

77. *Moore v. Publicis Groupe SA*, No. 11-CV-1279(ALC)(AJP), 2012 WL 1446534, at *3 (S.D.N.Y. Apr. 26, 2012).

78. *Id.* at *2.

79. *Id.*

80. *Global Aerospace Inc. v. Landow Aviation, LP*, No. CL 61040 (Va. Cir. Ct. Apr. 23, 2012) (Order Approving the Use of Predictive Coding for Discovery).

81. Memorandum in Support of Motion for Protective Order Approving the Use of Predictive Coding at 3, *Global Aerospace Inc.*, No. CL61040 (Va. Cir. Ct. Apr. 9, 2012).

82. *Id.* at 4.

records were unrelated to the project to construct the hangars that were at issue in this case. As a result, Landow found it necessary to cull substantially the data to generate smaller sets of documents that were potentially relevant and to avoid reviewing records that had nothing to do with the matter.

Defendant analyzed its options for review, which included linear review, culling by keywords, or using predictive coding, and determined that the use of TAR help reduce the expense of and time of a first-pass review.⁸³ They proposed to plaintiffs a predictive coding protocol.⁸⁴ The protocol provided that plaintiffs would receive a full set of training documents once the software was stabilized, but before defendant began iterations to separate the relevant from the irrelevant. The only records that would not be provided would be privileged records and sensitive records that would be tagged as irrelevant. Any withheld records would be logged sufficiently for opposing counsel to determine whether there was a need for further review of the coding decision. If plaintiffs' counsel believed that the documents were improperly coded, they could seek modification from the court before the software ran its iterations. Once the software had completed its iterations, defendant would conduct a statistically valid sampling program to establish that the majority of relevant documents were retrieved. They protocol set forth a recall level of 75% (based on that linear review averages recall of 59.3%).

Despite defendant's proposal, plaintiffs would not consent to the use of TAR, and instead they desired the use of keyword searching and linear review.⁸⁵ As a result, defendant filed a motion for a protective order permitting the use of TAR, arguing that plaintiffs proposed keywords were over-inclusive, and would produce numerous false positives.⁸⁶ Further, Defendant argued that keyword searching would likely retrieve only 20% of emails, would cause Landow to incur substantial review costs, and would require a substantial amount of time to review all of the data.⁸⁷ Unlike keywords, predictive coding would recover upwards of 75% of the records in substantially less time than keyword searching and linear

83. *Id.* at 5-6.

84. *Id.* at 11-12.

85. *Id.* at 13.

86. *Id.* at 8-9.

87. *Id.* at 8.

review.⁸⁸

Plaintiffs filed their objection arguing for the use of keyword searching.⁸⁹ Plaintiffs' counsel argued that Landow had a responsibility to produce *all* responsive documents located within their reasonable inquiry, not merely 75% percent as proposed in their ESI protocol. While plaintiffs agreed that defendant could use computer technologies to make the review process more efficient, computerized tools were supplements to an ordinary review.⁹⁰ "No computer program is an adequate substitute for having human beings review and sort the documents."⁹¹ The crux of plaintiffs' argument appeared to be that human review would be able to find all of the records, whereas TAR would not.

Defendant ultimately prevailed. Judge Chamberlain approved the use of predictive coding.⁹² The opinion did provide for plaintiffs to object after TAR had been utilized, should they feel the production was not complete, or to object to the ongoing use of predictive coding generally. Notwithstanding, not all courts are going to mandate the use of TAR, as was the case in *Kleen Products*.

C. *Kleen Products v. Packaging Corporation of America*

In April 2011, *Kleen Products, LLC. v. Packaging Corporation of America* also addressed the use of TAR.⁹³ This antitrust action is a set of consolidated cases.⁹⁴ Defendants are all manufacturers of containerboard, which is used by other manufacturers—in this case the plaintiffs—as lining material in other products, such as corrugated boxes.⁹⁵ In this dispute, plaintiffs alleged that defendants engaged in conscious parallelism, or price-fixing, in violation of the Sherman

88. *Id.* at 12.

89. Opposition of Plaintiffs: M.I.C. Industries, Inc., Factory Mutual Insurance Co., Global Aerospace, Inc., & BAE Systems Survivability Systems, LLC to the Landow Defendants' Motion for Protective Order Regarding Electronic Documents and "Predictive Coding" at 2, *Global Aerospace Inc. v. Landow Aviation, LP*, No. CL61040 (Va. Cir. Ct. Apr. 16, 2012).

90. *Id.* at 4.

91. *Id.*

92. *Global Aerospace Inc. v. Landow Aviation, LP*, No. CL 61040 (Va. Cir. Ct. Apr. 23, 2012) (Order Approving the Use of Predictive Coding for Discovery).

93. *Kleen Prods. LLC v. Packaging Corp. of Am.*, No. 10-C-5711, 2012 WL 4498465 (N.D. Ill. Sept. 28, 2012).

94. *Id.* at *1.

95. *Id.* at *1-3. *See also Kleen Prods., LLC v. Packaging Corp. of Am.*, 775 F. Supp. 2d 1071 (N.D. Ill. 2011).

Act.⁹⁶ Stated otherwise, defendants conspired to match their anticompetitive business behavior in setting prices on containerboard to the detriment of plaintiffs.

After records had been collected and searched using keywords, and the manual review had begun, plaintiffs requested that defendants change methods and use TAR to facilitate the review.⁹⁷ Plaintiffs believed that TAR would have been more accurate than using keyword searching and manual review. In December of 2011 the court referred discovery to Seventh Circuit Magistrate Judge Nan Nolan to help the parties resolve issues with search methods, search scope, and relevant time periods.⁹⁸ Hearings were held at which seven witnesses testified regarding the use of TAR.

In August 2012, the parties and the court issued a stipulation and order setting forth the parameters of electronic discovery.⁹⁹ Initially, during one of the hearings, Judge Nolan asked the parties to work together to reach agreement on the discovery issues, urging the parties to refine their keyword searches.¹⁰⁰ She noted in one of the hearings that she is a believer in the sixth Sedona Principle; that “the people who are producing records, producing the documents, are in a better position to know [what strategy to use].”¹⁰¹ In August, the parties set forth in a stipulation that plaintiff would not oppose to defendants’ method of production for the first round of review.¹⁰² The stipulation reserve the right of plaintiffs to seek the use of predictive coding should there be a second round of collection and discovery. Thus, despite plaintiffs’ efforts to require defendants to use predictive coding, at least for the first round of review, defendants retained using keyword searching and manual review to sift their materials.

96. *Kleen Prods., LLC*, 775 F. Supp. 2d at 1077.

97. See Joint Status Conference, *Kleen Prods. LLC v. Packaging Corp. of Am.*, No. 1:10-CV-05711 (N.D. Ill. May 17, 2012).

98. *Id.* at 2.

99. See Stipulation and Order Relating to ESI Search, *Kleen Prods. LLC*, No. 1:10-cv-05711 (N.D. Ill. Aug. 21, 2012).

100. *Id.* at 2-3; Transcript of Proceedings—Evidentiary Hearing before the Honorable Magistrate Judge Nan R. Nolan, Volume 2-A, at 299-300, *Kleen Prods., LLC*, No. 10 C 5711 (N.D. Ill. Mar. 28, 2012).

101. Transcript of Proceedings, *supra* note 100, at 297-98; *Sedona Conference*, *supra* note 12, at 193 (defining Principle 6 of The Sedona Principles).

102. See Stipulation and Order Relating to ESI Search, *Kleen Prods. LLC*, No. 1:10-cv-05711 (N.D. Ill. Aug. 21, 2012).

D. In re Actos (Pioglitazone) Products Liability Litigation

In re Actos addressed the use of TAR in late July, 2012, in the Western District of Louisiana.¹⁰³ This consolidated dispute arose from numerous claims, alleging that defendant's prescription drug for type 2 diabetes, called Actos, increased the risk of bladder cancer.¹⁰⁴ In addition, the claims alleged that defendants concealed their knowledge of this risk and failed to provide adequate warnings. Due to the large number of claims—potentially up to 10,000 claims spread across the country—the cases were consolidated into a multidistrict litigation in Louisiana, with U.S. District Judge Rebecca Doherty presiding.¹⁰⁵

On July 27, 2012, after some debate by the parties, Judge Rebecca F. Doherty issued a Case Management order setting forth the parties' protocol for using TAR for the production of defendant Takeda's ESI.¹⁰⁶ The detailed protocol, similar to the *Moore* protocol, required the parties to mutually identify custodians and work together to ensure that the test records were properly reviewed and tagged.¹⁰⁷ On numerous occasions, the protocol stated that the parties must work collaboratively. Moreover, the protocol allowed the plaintiff's experts to review both the not-privileged relevant and non-relevant records used to train the program, as well as to review a sampling of the records withheld as non-relevant from production.¹⁰⁸

The protocol provided details covering many aspects of the procedures, especially the assessment phase in, which the control set is determined.¹⁰⁹ The protocol outlined details on what percentages should be used to estimate the richness and, ultimately, the recall and precision of the dataset. In contrast with *Da Silva*, the protocol did not specify a cut off for the number of responsive records.¹¹⁰ Nor does the protocol address statistical sampling.

103. *In re Actos Prods. Liability Litig.*, 840 F. Supp. 2d 1356 (J.P.M.L. 2011).

104. Jef Feeley & Margaret Cronin Fisk, *Takeda Actos Judge Names Lead Attorneys in Cancer Suits*, BLOOMBERG (Mar. 13, 2012, 1:24 PM), <http://www.bloomberg.com/news/2012-03-13/takeda-actos-judge-names-lead-attorneys-in-u-s-bladder-cancer-lawsuits.html>.

105. *Id.*

106. *In re Actos (Pioglitazone) Prods. Liability Litig.*, No. 6:11-md-2299, 2012 WL 6061973 (W.D. La. July 27, 2012) (Case Management Order).

107. *Id.* at 6-16.

108. *Id.* at 7-8.

109. *Id.* at 10.

110. *Id.* at 8-9, 13.

E. EORHB, Inc. v. HOA Holdings, LLC

The most recent case to address the use of TAR is *EORHB, Inc. v. HOA Holdings, LLC*.¹¹¹ Unlike the preceding cases, Vice Chancellor J. Travis Laster *sua sponte* required the parties to use TAR, and even required them to use the same vendor.¹¹² The dispute arose from a commercial indemnity dispute involving the sale of the restaurant chain, Hooters.¹¹³ Judge Laster, towards the end of the hearing for summary judgment motions, ordered the parties to use TAR, noting that this was the type of case where “the parties would benefit from using predictive coding.”¹¹⁴ If any party disagreed, that party would need to show cause as to why predictive coding would not be suitable.¹¹⁵

The parties had not made any formal request for the use of TAR. Nor did the parties bring up the topic during the hearing.¹¹⁶ Rather, Judge Laster noted that “these types of indemnification claims can generate huge amounts of documents. That’s why I would really encourage you all, instead of burning lots of hours with people reviewing, it seems to me this is the type of non-expedited case where we could all benefit from some new technology use.”¹¹⁷

As of the drafting of this article, there has not been a motion brought by either party to oppose the order. Nor is there a protocol for the implementing TAR. However, Judge Laster’s order makes it clear that many courts, as well as parties, are becoming increasingly interested in using TAR to eliminate the discovery burdens posed by vast amounts of data.

IV. REMOVING THE FEAR FROM TECHNOLOGY ASSISTED REVIEW: CLARIFYING BEST PRACTICES AND GUIDELINES

Four courts have now specifically addressed the use of TAR. Yet the cases reveal the disagreement and uncertainty that exists for

111. *EORHB, Inc. v. HOA Holdings, LLC*, No. 7409-VCL (Del. Ch. filed Apr. 09, 2012)

112. Motion for Partial Summary Judgment, Motion to Dismiss Counterclaim and Ruling of the Court at 66, *EORHB, Inc.*, No. 7409-VCL (Oct. 15, 2012) (hearing transcript).

113. Ralph Losey, *NEWS FLASH: Surprise Ruling by Delaware Judge Orders Both Sides to Use Predictive Coding*, E-DISCOVERY TEAM (Oct. 25, 2012, 4:51 PM), <http://e-discoveryteam.com/2012/10/25/news-flash-surprise-ruling-by-delaware-judge-orders-both-sides-to-use-predictive-coding/>.

114. Motion for Partial Summary Judgment, *supra* note 112, at 66.

115. *Id.*

116. *See id.* at 66-67.

117. *Id.* at 67.

determining when and how TAR should be used. As TAR best practices, guidelines, and general use are better defined, legal disputes such as these will benefit. Clarifying standards will assist parties in reaching agreement earlier in the case and more easily, allowing for discovery to proceed more smoothly with less court interference.¹¹⁸

Parties will first need to determine whether TAR should be used at all. TAR is not suitable for all matters, and different tools will have different advantages and disadvantages depending on the nature of the case and the types of data. For example, matters with small amounts of data, or data with a lot of numerical information, images, heavy redaction requirements, short text-type messages, or complex coding schemes will not work as well with TAR, due to the nature of the data.¹¹⁹ On the other hand, matters with tight deadlines, textually dense materials, and substantial amounts of data will find TAR more beneficial.¹²⁰ Ultimately, whether the use of TAR would be beneficial will depend greatly upon the parties, the type of matter, the timeline, and the costs involved.

Cost is frequently a big factor in deciding whether to use TAR.¹²¹ Plaintiffs in *Moore* argued that the court and defendant placed too much reliance on costs in determining whether TAR should be used.¹²² Judge Peck acknowledged that, costs are indeed a factor in determining its use.¹²³ But costs, like other balancing factors, are frequently weighed against the other needs of the case. The argument of costs was also prevalent in *Global Aerospace*, where defendant argued that TAR was necessary because the size of the ESI was too costly to perform a traditional linear review.¹²⁴ If TAR is able to reduce the costs and time needed in comparison to traditional reviews, and it is as reliable as keyword searching and linear review, it should be considered.

If the parties elect to proceed with TAR, flexible guidelines and best practices that are set forth clearly will aid the parties in deciding how TAR will work best with their case. Cases are all different, with different data and different needs: production timelines differ; the amount and types of data differ; and, the sensitivity of the records and

118. Baron, *supra* note 73, at 244-45.

119. See, e.g., KERSHAW & HOWIE, *supra* note 27, at 28.

120. See *id.*

121. See generally *Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012).

122. Plaintiff's Rule 72(a) Objection, *supra* note 70, at 17-18.

123. *Moore*, 287 F.R.D. at 192.

124. Memorandum in Support of Motion for Protective Order, *supra* note 81, at 13-14.

the presence of privileged materials differ. Guidelines that are expressed rigidly may prevent parties from reaching agreement early in the case or from resolving discovery disputes as the case progresses. Rather, clear but flexible guidelines and best practices will aid parties in adapting TAR to cases' needs. "The people who are producing records, producing the documents, are in a better position to know [what strategy to use]."¹²⁵

Standards that act like rules, specifying precise measurements or requiring strict protocols that a case must meet, will not work with all matters. For example, in some cases, the parties may not want to or be able to meet a threshold requirement that the results of a review must meet: an 85% recall or an 85% precision rate. The types of data in the matter, how the data was stored, or the amount of data would not allow for that high of a return. Nor will some cases need to meet certain thresholds. For these cases, the parties may be less concerned with having a high recall level, and are instead more concerned with looking for specific materials. Rather than inquiring as to whether specific requirements were met, inquiries should focus on whether the parties' actions were reasonable and whether the relevant data was turned over.

Moreover, the Federal Rules of Civil Procedure do not require perfection, or production of *all* of the relevant evidence.¹²⁶ Rather, the parties to a dispute must act reasonably and in good faith to produce ESI.¹²⁷ "Counsel [must ensure] that a sufficient search has been conducted to obtain documents relevant to the case from which thorough discovery responses may be prepared."¹²⁸ Limits can be placed around discovery where the burden or expense outweighs the likely benefit, the needs of the case, the amount in controversy, the parties' resources, the importance of the issues at stake, and the importance of discovery for resolving the disputes.¹²⁹

In *Global Aerospace*, plaintiffs, who sought to use keyword

125. Transcript of Proceedings, *supra* note 100, at 297-98.

126. See *Moore*, 287 F.R.D. at 191-92; Paul & Baron, *supra* note 13, at 24. See also Pension Comm. of the Univ. of Montreal Pension Plan v. Banc of Am. Sec., LLC, 685 F. Supp. 2d 456, 461 (S.D.N.Y. 2010) ("In an era where vast amounts of electronic information is available for review, discovery in certain cases has become increasingly complex and expensive. Courts cannot and do not expect that any party can meet a standard of perfection."), *abrogated* by *Chin v. Port Auth. of N.Y. & N.J.*, 685 F.3d 135 (2d Cir. 2012).

127. See generally *Sedona Conference*, *supra* note 12. See also *Moore*, 287 F.R.D. at 192.

128. Memorandum in Support of Motion for Protective Order, *supra* note 81, at 18.

129. *Moore*, 287 F.R.D. at 192 (citing FED. R. CIV. P. 26(b)(2)(C)).

searching and linear review, objected to defendant's request to have a 75% measure of recall using TAR, arguing that this would produce incomplete results.¹³⁰ However, even if the parties used a manual review, there would be no guarantee that 100% of the relevant data would be produced: humans are not perfect. In fact, research shows that errors in manual review are likely to result in substantially less than a 100% recall rate. In many cases, this percentage has been shown to be less than 50%.¹³¹ No search tool—keywords with manual review or otherwise—can guarantee that all responsive documents will be identified. A 100% recall rate is just not realistic.¹³²

This notion of an imperfect return of all relevant records seems to have been missed by many attorneys.¹³³ In fact, focusing on the completeness of the return is oftentimes less important than focusing on what is actually returned. Consider that in one case, a 100% recall rate may produce 95% relevant emails that add nothing material to the case and only 5% that add true substantive value. In another with substantially different data, 95% of the relevant emails may add material value and only 5% do not. Being able to determine what balance will work best will depend on an analysis of the records, the needs of the case, and the resources available to the parties.¹³⁴

Flexible standards will also allow parties to properly negotiate their ESI protocols so that the requirements meet the needs of the case. *Moore, Global Aerospace*, and *In re Actos* provided fairly detailed protocols setting forth how TAR would be used in the matters. The 22-page *Moore* protocol, for example, set forth particular requirements for measuring the reliability of the process: the numbers of records used to create samples and seed sets, the number of potential iterations needed to complete the process with built in flexibility, and the necessary percentages needed for satisfaction.¹³⁵ It also set forth a detailed search methodology, such as how the initial seed sets would be collected and reviewed, what tagging would be used, and what materials defendant would provide to plaintiffs.¹³⁶ But the *Moore* requirements, even having built-in flexibility for this

130. Opposition of Plaintiffs to the Defendants' Motion, *supra* note 89, at 2.

131. See, e.g., Blair & Maron, *supra* note 22, at 293 (finding that recall was only 20% of the relevant documents).

132. *Sedona Conference*, *supra* note 12, at 211.

133. Oot et al., *supra* note 30, at 545-46.

134. *Sedona Conference*, *supra* note 12, at 209-11.

135. *Moore v. Publicis Groupe*, 287 F.R.D. 182, 201-03 (S.D.N.Y. 2012).

136. *Id.* at 199-200.

matter, would not necessarily be suitable for another matter.

As part of the process for using TAR, parties will need to select some form of quality verification.¹³⁷ Some forms of quality verification methods include using independent testing to replicate and confirm results, adoption of reconciliation reports, and simply inspecting the data to verify and report discrepancies.¹³⁸ One common and well known form of quality verification is sampling, though its application to discovery is only a recent phenomenon.¹³⁹ For example, in keyword searching it has been rare for a party to sample records removed as nonresponsive by the searches.¹⁴⁰ Thus, parties may never have deduced whether the search terms were sufficiently inclusive; there could have been numerous records not captured by incorrect or narrow search terms.

Large pools of ESI make sampling a necessity to satisfy a reasonableness standard of production. Sampling furthers the desirability of TAR by aiding in confirming a party's results and satisfying the other parties of the contents of the production.¹⁴¹ In essence, sampling supports defensibility. Courts recognize this. Due to problems created by a lack of sampling—producing numerous non-relevant records or failing to produce relevant records—courts have increasingly disfavored parties that failed to sample and verify results of keyword searches.¹⁴²

As with setting recall and precision rates, setting a precise number or percentage of records to sample makes little sense.¹⁴³ Sampling is not a one-size-fits-all verification method.¹⁴⁴ How large of a sample is needed and what statistical confidence level will be used will depend on the parties' choices and what is being

137. Quality verification is also commonly referred to as quality control, or QC. *See* THE SEDONA CONFERENCE, COMMENTARY ON ACHIEVING QUALITY IN THE E-DISCOVERY PROCESS 9-10 (2009) [hereinafter SEDONA CONFERENCE COMMENTARY]; Baron, *supra* note 27, at 29-33; Conor R. Crowley, *Defending the Use of Analytical Software in Civil Discovery*, 10 DIGITAL DISCOVERY & E-EVIDENCE no. 16, Sept. 16, 2010.

138. Baron, *supra* note 27, at 30.

139. *See, e.g.*, KERSHAW & HOWIE, *supra* note 27, at 17.

140. *See* Memorandum in Support of Motion for Protective Order, *supra* note 81, at 18.

141. KERSHAW & HOWIE, *supra* note 27, at 17.

142. Oot et al., *supra* note 30, at 553-54. *See also* William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 136 (S.D.N.Y. 2009) (noting that electronic discovery requires cooperation and transparency); *In re Seroquel Prods. Liability Litig.*, 244 F.R.D. 650, 662 (M.D. Fla. 2007).

143. SEDONA CONFERENCE COMMENTARY, *supra* note 137, at 11.

144. *See id.* at 13; Paul & Baron, *supra* note 13, at 30.

evaluated.¹⁴⁵ For example, sampling privileged materials may require a higher confidence level.¹⁴⁶ The same may apply for more complex data or for data where the reviewers tagging it were not as knowledgeable about the case. Data that has more sensitive records may need to be screened more thoroughly and with more restrictive parameters. The pace of the case may impact how much sampling is done and when it is done as well. Regardless, sampling should be used and tailored to fit the needs of the case.

Quality verification as well as compliance with any other standards requires collaboration and transparency.¹⁴⁷ To collaborate, the parties will need to interact with one another on a frequent basis throughout the discovery process.¹⁴⁸ Parties should make a good faith attempt to discuss the use of particular search and information retrieval methods, tools, and protocols, including whether TAR should be used.¹⁴⁹ The parties will need to share samples of responsive ESI, discuss search terms, and iteratively refine the search protocol to reach a consensus.¹⁵⁰ Having a senior attorney among the leadership be present to ensure quality verification methods are utilized satisfactorily will help further the parties' objectives.¹⁵¹ If no agreement for quality verification is obtained, a party can either abandon TAR's use or attempt to obtain approval through the courts.¹⁵²

In order to facilitate collaboration, the parties will also need to be transparent.¹⁵³ Parties will want to utilize TAR in a public and transparent manner and be open to the other parties from the outset.¹⁵⁴

145. See KERSHAW & HOWIE, *supra* note 27, at 17.

146. See *id.*

147. See *Moore v. Publicis Groupe*, 287 F.R.D. 182, 192 (S.D.N.Y. 2012); *In re Seroquel Prods. Liability Litig.*, 244 F.R.D. at 662; Baron, *supra* note 27, at 30.

148. See, e.g., *In re Actos (Pioglitazone) Prods. Liability Litig.*, No. 6:11-md-2299, 2012 WL 6061973 (W.D. La. July 27, 2012) (stating that the parties should "collaborate" throughout the protocol).

149. *William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134, 136 (S.D.N.Y. 2009).

150. Baron, *supra* note 27, at 30. See also *Moore*, 287 F.R.D. 182. Note that while the defendants in *Moore* set forth in the protocol that they would share both samples of records that were coded as responsive as well as those that were coded as not responsive, a party to a suit has no legal requirement to provide another party non-responsive records. All that is required under the Federal Rules is for a party to provide responsive, non-privileged data.

151. See Baron, *supra* note 27, at 31; *Sedona Conference*, *supra* note 12, at 199.

152. Peck, *supra* note 1.

153. *Moore*, 287 F.R.D. at 192.

154. *William A. Gross Constr. Assocs., Inc.*, 256 F.R.D. at 136; PACE & ZAKARAS, *supra*

Courts will look more favorably upon a party who discloses its key custodians and how it will for the requested documents. Where a party is transparent, “opposing counsel and the Court are more apt to agree to your approach”¹⁵⁵

Thorough documentation of the steps taken is necessary.¹⁵⁶ Proof of valid testing should be recorded.¹⁵⁷ Most TAR tools have the ability to audit the process and provide results of the review.¹⁵⁸ In fact, reporting of audits should be a required feature of the tool selected by the parties. The parties should take extensive notes about what was done, when it was done, and what the findings were. Courts look more favorably upon parties who have documented their steps and can explain what they did than they do upon those who failed to document their procedures and are unable to explain the methodology chosen.¹⁵⁹

Standards guiding TAR’s use are only a start, however. Attorneys need to be more knowledgeable about technology and be involved in the entire e-discovery process, not just the substantive review.¹⁶⁰ “[J]ust as important as utilizing the automated tools, is tuning the *process* in and by which a legal team uses such tools”¹⁶¹ A well-thought out process with substantial human input will improve the chances of succeeding in using any automated search method.¹⁶² Lawyers cannot assume that they can make a quick and simple decision about how to use TAR, let alone any other search and review method.¹⁶³ The chosen process, and especially the process with TAR, requires iterative decision-making, which allows for feedback and learning and ultimately measurement of results.¹⁶⁴ “The

note 31, at 81.

155. *Moore*, 287 F.R.D. at 193. *See also William A. Gross Constr. Assocs., Inc.*, 256 F.R.D. at 136; *In re Seroquel Prods. Liability Litig.*, 244 F.R.D. 650, 662 (M.D. Fla. 2007).

156. *Sedona Conference*, *supra* note 12, at 199.

157. *Peck*, *supra* note 1.

158. *KERSHAW & HOWIE*, *supra* note 27, at 21.

159. *See Victor Stanley, Inc. v. Creative Pipe, Inc.*, 250 F.R.D. 251, 262 (D. Md. 2008); *Peck*, *supra* note 1. *See also William A. Gross Constr. Assocs., Inc.*, 256 F.R.D. at 136.

160. William W. Belt, Dennis R. Kiker & Daryl E. Shetterly, *Technology-Assisted Document Review: Is It Defensible?*, 18 RICH. J.L. & TECH. no. 3, at 1, 25 (2012) (“Technology is only reliable when it is used in conjunction with the right process.”).

161. *Sedona Conference*, *supra* note 12, at 199.

162. *Id.* at 199, 209.

163. *Victor Stanley, Inc.*, 250 F.R.D. at 262 (commenting that search and retrieval methods require the utmost care in selecting methodology appropriate for the task and that failure to do so may result in disclosure of privilege).

164. *Sedona Conference*, *supra* note 12, at 199.

time and effort spent on the front end designing a sophisticated discovery process that targets the real needs of the client must be viewed as a condition precedent to deploying automated methods of search and retrieval.”¹⁶⁵

Additionally, the attorneys making the relevancy designations should be well-versed in the issues of the case.¹⁶⁶ Analytical software requires reviewing attorneys to review records based on the concepts in the documents. A review based on concepts results in a higher level of recall.¹⁶⁷ But an understanding of the concepts requires an understanding of the case. Ultimately, determinations made as a result of a sound understanding of the case lead to relevancy determinations being made by the system from the best available data. The results will be more accurate for the needs of the case.

Automated search methods may be more reasonable and valuable compared with manual review, and, in some cases, even necessary.¹⁶⁸ The choice of a method will depend on the legal context in which it will be deployed.

The needs of the litigation at issue should dictate what technology gets used and how, in order to strike the optimal balance possible between recall and precision. The key to defensibility is that litigants employ these search strategies as part of a reasonable, good-faith, well-documented discovery protocol. “Lawyers must understand where the search technology fits into that protocol and have confidence that they have taken measures to ensure the quality of their searches.”¹⁶⁹

Ultimately, what software is used is less important than the methodology used to optimize the review using TAR.¹⁷⁰ The methodology used will be strongly aided by standards that are flexible and can guide attorneys on how to fit TAR to their disputes.

165. *Id.*

166. Crowley, *supra* note 137.

167. *Id.*

168. *Sedona Conference*, *supra* note 12, at 194 (stating Practice Point 1).

169. Mia Mazza, Emmalena K. Quesada & Ashley L. Sternberg, *In Pursuit of FRCP 1: Creative Approaches to Cutting and Shifting the Costs of Discovery of Electronically Stored Information*, 13 RICH. J.L. & TECH. no. 3, at 1, 34 (2007).

170. Amanda Jones & Ben Kerschberg, *What Technology-Assisted Electronic Discovery Teaches Us about the Role of Humans in Technology*, FORBES (Jan. 9, 2012, 10:18 AM), <http://www.forbes.com/sites/benkerschberg/2012/01/09/what-technology-assisted-electronic-discovery-teaches-us-about-the-role-of-humans-in-technology/>.

CONCLUSION

Using TAR can greatly benefit cases, but its use should be evaluated on a case-by-case basis. Flexible standards and best practices will help guide parties in determining when and how to use TAR. Where a case has a sufficient amount of textual data, TAR may very well reduce costs and the time for preparing the data. But parties to an action must be circumspect in deciding how they will use TAR. They must be prepared to collaborate with each other and be open about how they plan to use TAR. Thorough sampling must be conducted and extensive note-taking must be performed to improve defensibility. The parties must evaluate their data both before and during the use of TAR to find recall and precision rates that are appropriate for the cases. Lastly, attorneys must spend the time to master the technology; a failure to understand the technology will increase risks of negative rulings and court sanctions.

The e-discovery rulings that have been handed down over the past year helped facilitate the use of technology assisted review. Each court provided crucial decisions and testimony surrounding its use helping to guide parties' expectations.¹⁷¹ With time—and as attorneys and the legal profession become more comfortable with alternative technologies—more and more parties are likely to embrace the benefits of TAR when the case would be benefitted by it. As it is used, and as more precedent and commentary are developed, what has worked and what does not will be more clearly set forth. And with that, hopefully, there will be fewer disputes and an ability to move through discovery more smoothly, so that parties can get to the heart of the dispute instead of wasting money and time arguing about logistics.

171. See, e.g., Motion for Partial Summary Judgment, *supra* note 112 (*sua sponte* requiring the parties to use TAR).