

Technology-Assisted Review for Discovery Requests

A Pocket Guide for Judges

Timothy T. Lau

Emery G. Lee III

Federal Judicial Center
2017

This Federal Judicial Center publication was undertaken in furtherance of the Center's statutory mission to conduct and stimulate research and development for the improvement of judicial administration. While the Center regards the content as responsible and valuable, this guide does not reflect policy or recommendations of the Board of the Federal Judicial Center.

Contents

Introduction, 1

How TAR Differs from Other Review Methods, 2

Sufficiency of TAR Under Federal Rule of Civil Procedure 26(g), 6

When the Use of TAR Meets the Requirements of Federal Rule of Civil Procedure 26(g), 6

When the Conduct of TAR Meets the Requirements of Federal Rule of Civil Procedure 26(g), 8

 Proper Seed Set Construction, 9

 Statistical Validation, 11

Appendix: An Illustrative TAR Order, 13

For Further Reference, 17

Introduction

Discovery is especially costly in civil cases that involve reviewing a large number of electronic documents for relevance and privilege. Recent developments in information retrieval technology have enabled the use of computers to actively assist in document review, promising to help reduce the overall cost of discovery. The first judicial opinion approving the use of technology-assisted review (TAR) was issued in 2012,¹ and the use of TAR has since become more common. Though this review method goes by many names (e.g., computer-assisted review, predictive coding), we will refer to it as TAR.

This pocket guide is intended to provide judges and court staff with a brief, understandable introduction to TAR and the issues that may arise when it is used. As with every solution for a problem, the devil is in the details. TAR is not a one-size-fits-all solution to discovery issues. It is also a new and developing technology. There are many vendors in this space, whose methods vary and will continue to evolve. For these reasons, this guide provides an overview and some general points, but it does not delve into the specifics of how particular TAR tools work. Of course, because TAR tools differ, not everything in this pocket guide is applicable to every tool.

TAR may be complex and difficult, but standard case-management strategies are still effective. Judges managing complex cases involving TAR should take a proactive approach to address the issues discussed in this pocket guide early in the case, preferably at the first scheduling conference, requiring the parties to negotiate a workable plan for discovery. Once discovery begins, the parties should keep the court apprised of the progress of that plan.

Judges should require the parties to educate them on the details of the process, preferably before it begins. A knowledgeable representative from the TAR vendor assisting the party may be helpful.

While TAR and associated technologies are new, the flexibility inherent in the Federal Rules of Civil Procedure, which empowers judges to guide cases toward a just, speedy, and (relatively) inexpensive resolution, allows judges sufficient authority to direct the use of TAR techniques in those cases where these techniques may be most useful.

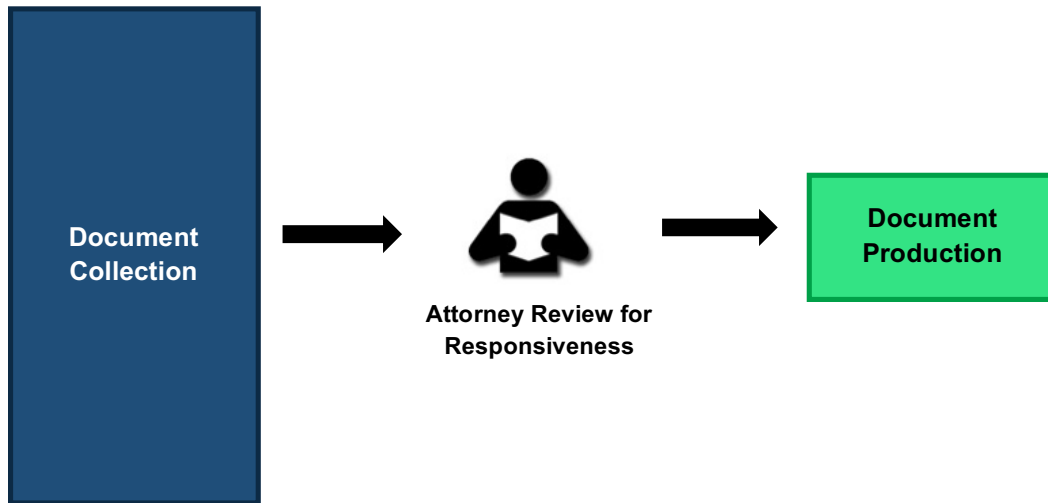
After a brief description of how TAR differs from more traditional forms of review, this guide addresses two particularly vexing case-management issues that arise in TAR cases: the disclosure of documents and the adequacy of TAR in meeting the “reasonable inquiry” requirement under Federal Rule of Civil Procedure 26(g). To these ends, an illustrative order for the use of TAR is included in the appendix. A list of additional resources can be found at the back of the guide.

1. *Da Silva Moore v. Publicis Groupe & MSL Grp.*, 287 F.R.D. 182 (S.D.N.Y. 2012), *aff'd*, No. 11 Civ. 1279 (ALC) (AJP), 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).

How TAR Differs from Other Review Methods

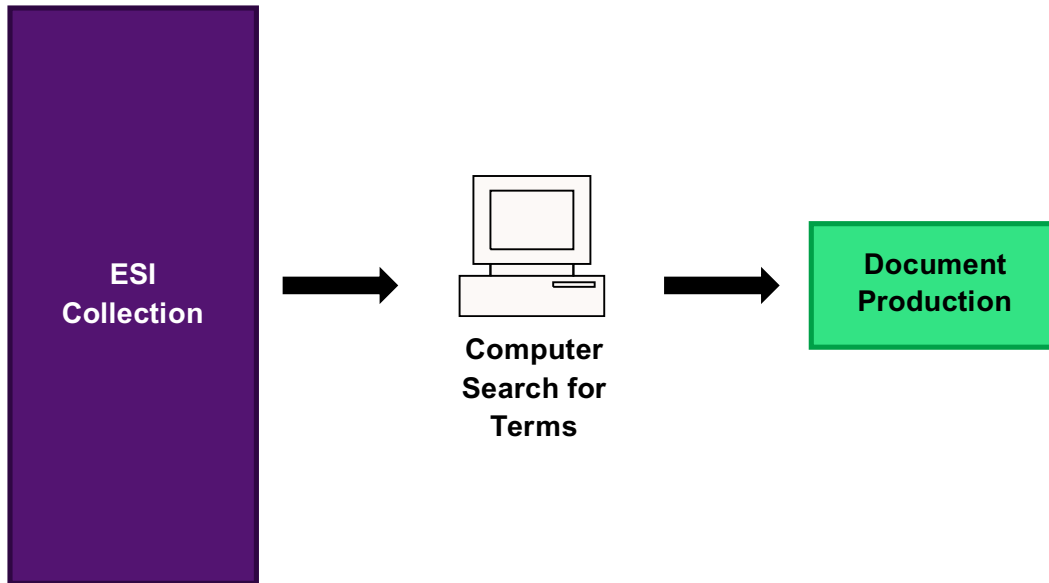
Under Federal Rule of Civil Procedure 26(b)(1), a party may obtain documents “relevant to [his or her] claim or defense” by propounding document requests. The responding party is then required to produce any non-privileged documents that are responsive to the requests.

Traditionally, the responding party has attorneys manually review all the documents in its collection to determine which documents to produce. This is sometimes referred to as manual or linear review. The non-privileged documents identified as responsive are produced.



The process is fundamentally subject to attorney subjectivity. Different attorneys reviewing the same document may reach different conclusions as to the document’s responsiveness. Also, with the proliferation of computer technologies, the number of business documents generated and stored has grown significantly. When a civil dispute arises, there is often far too much electronically stored information (ESI) retrieved than attorneys can practically review for responsive documents.

Computers are now commonly used to search collected ESI for search terms. Through a meet-and-confer process, parties generally agree upon the specific terms used to search the ESI collection, and the documents containing the search terms will be treated as responsive or will be passed on for further attorney review for responsiveness and privilege.



Search terms are convenient because practically all attorneys have personal experience with Internet search engines that identify responsive websites based in part on search terms. Once the parties have agreed to specific terms, there are also fewer grounds for disputes.

Legal questions often map directly onto specific search terms. For example, establishing contributory patent infringement or inducement under 35 U.S.C. § 271(b)–(c) requires proof of “knowledge of the existence of the patent that is infringed.”² A patent owner seeking to prove such knowledge can ask for documents that mention the specific patent number. A patent number is seven digits long, and given the uniqueness of such a number, it is reasonable to assume that any collected document containing that precise number is potentially responsive.

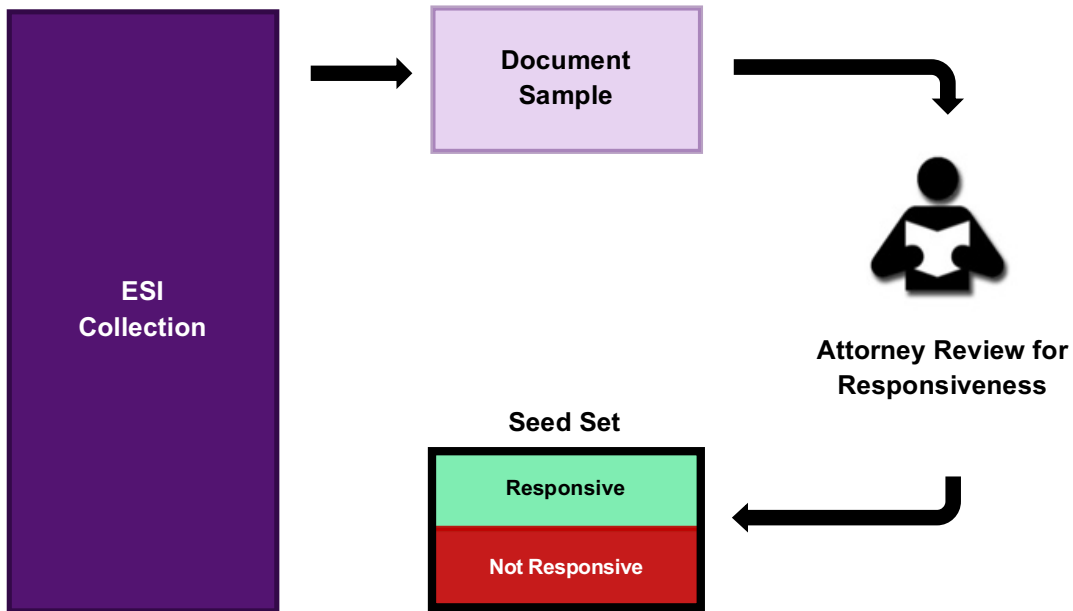
Not all legal questions can be boiled down to such pinpointed searches, however. “Price-fixing agreements between two or more competitors . . . fall into the category of arrangements that are *per se* unlawful.”³ One could propose search terms of *agree*, *fix*, and *price* to search for documents about price-fixing agreements, but unlike the aforementioned example of a patent number, these terms are likely to yield overinclusive results. Many documents may use one or more of these terms without bearing on price-fixing agreements. Moreover, the words *agree*, *fix*, and *price* all have multiple synonyms. Documents expressing an agreement to fix prices need not use any of these words. It is not always easy to translate a search request written for and understood by humans into a simple computer search command.

2. *Global-Tech Appliances, Inc. v. SEB S.A.*, 131 S. Ct. 2060, 2068 (2011).

3. *Texaco Inc. v. Dagher*, 547 U.S. 1, 5 (2006) (interpreting § 1 of the Sherman Act).

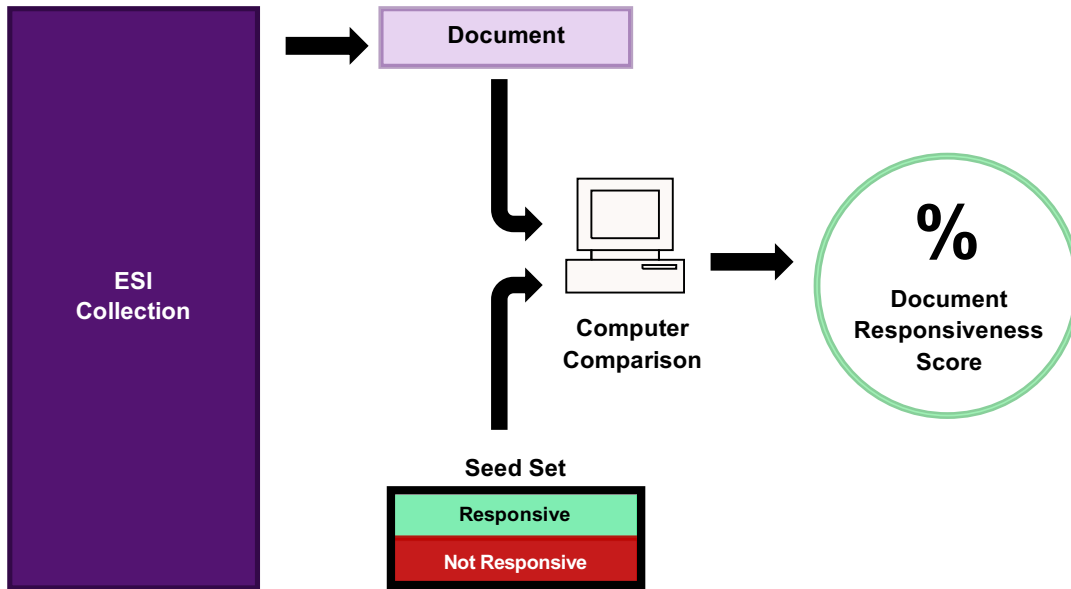
TAR seeks to avoid this problem by combining human review with computerized search. There are many variants of TAR, and the exact implementation differs from one vendor to another. Nonetheless, the core concept is to review an ESI collection by having a computer mimic the judgment of attorneys based on their review of a subset of the documents. The general technique runs as follows.

In the beginning, a sample of the ESI collection is reviewed by the attorneys of the responding party.⁴ This sample may be collected by a variety of means, including random sampling, judgmental sampling, and the use of search terms. The reviewing attorneys use their subject-matter expertise to determine which of the documents within the sample are responsive and which are not. This sample is referred to as the seed set.



Then the computer is put to work. Each document in the ESI collection is compared with the seed set, and the computer determines a responsiveness score for each one. Documents may be scored, for example, based on the similarity of their text to the text of responsive documents in the seed set. Documents that score above a chosen responsiveness threshold are then marked as responsive.

4. These attorneys are sometimes referred to as subject-matter experts.



To check the adequacy of the responsiveness scoring, the attorneys will review a sample of the documents that the computer tagged responsive or unresponsive, re-coding documents as appropriate. The computer then compares each document in the ESI collection again with all of the documents reviewed by the attorneys. Using the expanded set of attorney-reviewed documents for guidance, the computer will generate a new set of responsiveness scores. This process, called training, continues until the results are acceptable. The resulting documents are deemed responsive and produced after a privilege review.

TAR therefore eliminates the need to reduce determination of responsiveness to a single rule or a search term. Instead, assuming that documents similar to those the attorneys found to be responsive are also responsive, TAR can determine the responsiveness of all documents in an ESI collection by using the attorneys' determinations about a smaller set of documents. The potential cost savings of TAR come from replacing attorney review of most or all of the ESI documents with computer determinations that significantly reduce the number of documents attorneys must review.⁵

5. The cost of manual review of the documents deemed responsive by TAR for privilege issues is unavoidable.

Sufficiency of TAR Under Federal Rule of Civil Procedure 26(g)

Federal Rule of Civil Procedure 26(g) requires an attorney to certify that a discovery response is made “to the best of . . . knowledge, information, and belief formed after a reasonable inquiry.” A Rule 26(g) challenge to a production of documents selected by TAR is likely to raise one of two concerns:

1. whether, in general, the use of TAR to respond to a particular request constitutes a “reasonable inquiry” and
2. whether the TAR review, as it was conducted, constitutes a “reasonable inquiry.”

The following sections address some of the considerations relevant to each question.

When the Use of TAR Meets the Requirements of Federal Rule of Civil Procedure 26(g)

Because statistical approximation, not actual responsiveness, determines the number of documents to be produced in a TAR review, attorneys certifying TAR production know that a production most likely does not include all responsive documents and that it does include those that are not responsive. Courts have therefore been asked from time to time to decide whether using TAR to select documents for production constitutes a “reasonable inquiry.”

The fact that TAR will never yield perfect results does not mean that a TAR process cannot suffice as a “reasonable inquiry.” The Committee Notes to Rule 26 state that “[t]he duty to make a ‘reasonable inquiry’ is satisfied if the investigation undertaken by the attorney and the conclusions drawn therefrom are reasonable under the circumstances.” As noted in an often-cited opinion on TAR, “the Federal Rules of Civil Procedure do not require perfection.”⁶ In fact, research suggests that the traditional methods of manual document review and use of search terms to identify responsive documents also result in many missed documents—because of error in human judgment or underinclusive search terms.⁷

A key consideration, then, is whether a TAR process is technically appropriate for a particular request. TAR can be expected to perform particularly well in a review if the following is true:

1. The number of documents in a collection is large.

6. *Da Silva Moore*, 287 F.R.D. at 191.

7. See generally Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and Efficient Than Exhaustive and Manual Review*, 17 *Richmond J.L. & Tech.* 11 (2011).

2. Responsive documents are expected to be similar to each other in some fashion.⁸
3. The TAR algorithm measures that similarity.⁹

TAR is particularly useful for identifying documents if search criteria are too complex to be defined or specified.

For all its sophistication, TAR is not always better than search terms at identifying responsive documents. If unique identifiers or words are indicative of a document's responsiveness, the TAR process will not be likely to outperform the use of search terms. Judges may consider the relative advantage of TAR over search terms based on the nature of a particular document request.

Also, under the proportionality requirement of Rule 26, cost is a particularly important factor. As discussed earlier, alternative processes for determining responsiveness will sometimes be superior to TAR. But TAR is able to greatly reduce the number of documents in a collection that need to be reviewed by humans. The cost savings may, in some cases, make TAR a "reasonable inquiry," but cost must be balanced against other factors as well.¹⁰

Search terms have the comparative advantage of familiarity. Attorneys and judges are all accustomed to the art of proposing search terms, given their widespread use in everything from Internet search engines to online services for researching judicial opinions. This means the requesting party can have some pre-review participation in deciding how a particular review is conducted, since during the meet-and-confer process, it can propose and negotiate search terms it considers appropriate. If the parties fail to agree on search terms, the court can readily adjudicate a dispute and rule on the reasonableness of proposed terms.

In contrast, the TAR process may seem like a black box, incapable of being challenged. The evaluation of a seed set and of any particular TAR review process is complex. To give an example, the proper formation of a seed set requires it to be representative of the whole collection. But the construction of a representative sample of

8. One should not always assume that responsive documents are similar to each other. Take, for example, the ancient parable of the six blind men who, when asked to observe an elephant by touching, reported very different observations. The six statements may bear little similarity to each other in content; yet all six statements are responsive by any measure to the question about the nature of the animal.

9. Each TAR algorithm has its own definition of similarity, and any TAR algorithm used in a particular review should have a definition of similarity that corresponds to the similarities expected in responsive documents. If the responsive documents are similar to each other in language, then the TAR algorithm to be used should be one that compares language. However, if responsive documents are similar not in language but in their overall visual resemblance, one would need to use a TAR algorithm capable of weighing graphic similarity.

10. See Elle Byram, *The Collision of the Courts and Predictive Coding: Defining Best Practices and Guidelines in Predictive Coding for Electronic Discovery*, 29 Santa Clara Computer & High Tech. L.J. 675, 694 (2013).

documents is part science and part art, with many problems similar to those that arise in the construction of representative population samples for opinion polling. Just as the construction and criticism of sampling methodologies for opinion polling is a specialist subject matter, evaluating the adequacy of seed sets must at this time be considered beyond the skill of the average lawyer.

Also, many of the computer algorithms vendors use to calculate responsiveness scores during the TAR process are proprietary. These algorithms may take any number of factors into consideration, such as the time stamp of the document, the similarity of text, and the author of a document. But what these factors are, what weight each factor is given, and how each factor is used to weigh responsiveness differ from vendor to vendor. And even if vendors are willing or can be compelled to disclose their algorithms, the average attorney will be unable to evaluate such computer code.

Complicating the task of seed set and TAR algorithm evaluation is the fact that proper seed set construction is itself dependent on the TAR algorithm to be used. Some algorithms rely on seed sets formed by random sampling, while others may require judgmental sampling. Even if an attorney is familiar with the assumptions used in one vendor's TAR process, one cannot assume that he or she is able to evaluate the review process of another vendor. There is, to our knowledge, no published manual that evaluates the algorithms of each vendor and explains the requirements of each algorithm for proper seed set construction.

For these reasons, challenging a TAR production effectively is the province of the most sophisticated of requesting parties, who have the resources to afford the technical expertise. Should a dispute over TAR arise, a judge may find that he or she will require education from experts to understand the arguments of the parties. In deciding whether to permit the use of TAR over the objection of a requesting party, the judge may therefore take into consideration the party's ability to critique and challenge a defective review. In those circumstances where the use of TAR does not offer advantages over the use of search terms, traditional methods may be more reasonable.

When the Conduct of TAR Meets the Requirements of Federal Rule of Civil Procedure 26(g)

The Committee Notes to Rule 26 state that

[t]he duty to make a "reasonable inquiry" is satisfied if the investigation undertaken by the attorney and the conclusions drawn therefrom are reasonable under the circumstances. It is an objective standard similar to the one imposed by Rule 11. See the Advisory Committee Note to Rule 11. . . . Ultimately, what is reasonable is a matter for the court to decide on the totality of the circumstances.

Rule 26(g) does not require the signing attorney to certify the truthfulness of the client's factual responses to a discovery request. Rather, the signature certifies

that the lawyer has made a reasonable effort to assure that the client has provided all the information and documents available to him that are responsive to the discovery demand.

Because TAR is a process built on statistical principles, evaluating the performance of a TAR review is based on statistics and is likely to be extremely complex. Rule 26 itself provides no guidance on the necessary level of statistical tolerance. This section outlines some of the possible considerations and also suggests some case-management practices that may help prevent TAR-related disputes.¹¹

Proper Seed Set Construction

Mistakes and biases in the seed set will translate into mistakes and biases in the eventual document production. As discussed earlier, the attorneys' review of the seed set essentially serves as the template from which the TAR algorithm reviews all other collected documents. TAR therefore relies heavily upon a properly formulated seed set. Any error in attorney judgment during formation and review of the seed set will be built into the computer's determination of all the other documents' responsiveness. Any error in selecting the seed set documents will produce an inaccurate yardstick, corrupting the training process and thereby the results of a TAR review.

A challenge to a particular TAR review may focus on the attorney review or the construction of the seed set, but it is very difficult for the court to wade into a particular seed set to judge the statistical validity of its construction and the quality of the review. Instead, the court may want to rely on the requesting party or its expert to point to specific problems.

To that end, judges should encourage the parties to be transparent about the construction of the seed set. The responding party will of course include the responsive, non-privileged documents of the seed set when it produces all the other responsive, non-privileged documents. But if the responding party does not identify which ones are the seed set documents, the requesting party will have little knowledge on which to base a challenge to the seed set's constitution.

The predominant approach, as of this writing, is to have the responding party create the seed set by itself. In these circumstances, the responding party ideally agrees to disclose all the non-privileged documents in the seed set to the requesting party, whether or not they are responsive. Producing the full seed set allows the requesting party to review the documents, if it so chooses. If it finds potential problems in the review, it can ask the responding party to remedy the defects and possibly settle a brewing dispute without the intervention of the court. The requesting party can

11. To the best of our knowledge, the question whether a particular TAR review, as it was conducted, constitutes a "reasonable effort" under the "objective standard" articulated under Rule 26(g) has not been raised in a court.

also evaluate whether the seed set can statistically represent the entire document collection.

If it is not feasible for the responding party to produce the full seed set, it can at least identify which of the produced documents were part of the seed set. The requesting party can then look within those documents to get some idea of the criteria that the responding party used to determine responsiveness. Though without the entire seed set, the requesting party cannot tell whether the responding party tried to influence the TAR algorithm to conceal responsive documents by flagging only a select portion of actually responsive documents within the seed set as responsive,¹² the requesting party will at least know that the responding party did not hide the responsive documents that were similar to those select seed set documents.¹³

There have been cases where responding parties would not disclose the seed set to the requesting parties, and district judges have found they lack the power to compel such disclosures.¹⁴ This is a developing area of law, but under the current Federal Rules of Civil Procedure, judges may not be able to require responding parties to do more than produce relevant, non-privileged documents.

An alternative approach to avoiding disputes over the creation of the seed set might be described as the “experts in a locked room” approach, in which both producing and requesting parties designate subject-matter experts to construct the seed set together. These subject-matter experts meet in the same room to review documents for inclusion in the seed set. By court order, these experts are not to disclose or improperly use any non-responsive or privileged information obtained during this process. This approach has been attempted only once thus far in a recent case,¹⁵ and it is too early to conclude whether this approach offers practical benefits over the predominant approach.

In any case, judges should urge the parties to agree on a transparent, cooperative TAR process, just as they now encourage parties to discuss and agree on search terms

12. Suppose there are two classes of actually responsive documents in the collection, type A and type B, where the production of type B documents would be more damaging to the responding party. An attorney who wishes to conceal type B documents could, in the review of the seed set documents, mark only the type A documents as responsive. The TAR process, emulating the attorneys’ review of the seed set, would identify only type A documents in the entire collection as responsive and would identify type B documents as nonresponsive. The resulting production set would therefore be stripped of type B documents.

13. Returning to the example in note 12, by observing that type A documents among the seed set were marked as responsive, the requesting party can know that the TAR process would have picked up most of the type A documents within the collection, should it be functioning as claimed.

14. See John M. Facciola & Philip J. Favro, *Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled to Work Product Protection*, 8 Fed. Cts. L. Rev. (Feb. 2015).

15. *In re Bair Hugger Forced Air Warming Prod. Liab. Litig.*, No. 0:15-md-02666 (D. Minn. July 8, 2016), ECF No. 62, <http://www.mnd.uscourts.gov/MDL-Bair-Hugger/Orders/2016/2016-0708-PTO12-ECF-Stamped-Order.pdf>.

prior to a review that uses them. Again, ideally the parties should discuss and agree to the use and conduct of TAR early in the case, prior to the beginning of the review.

Statistical Validation

Judges should encourage responding parties to share how they determined that the TAR process produced an accurate result.

A well-constructed seed set is a necessary but insufficient condition for establishing a proper TAR process. The ultimate document production, which is built from the seed set, must itself undergo a statistical validation.

There are two generally accepted metrics for evaluating TAR document production.¹⁶ The first is precision, which is the fraction of documents tagged responsive by TAR that actually are.¹⁷ In short, it is a measure of the correctness of a TAR review. Precision can be calculated in many ways, such as by an attorney review of documents TAR identified as responsive. A low precision score means the produced documents are highly overinclusive.

Ideally, the precision should be 100%; that is, all documents that TAR identifies as responsive are actually responsive. However, as discussed before, TAR determines the responsiveness of collected documents by emulating the attorneys' review of the seed set; it does not seek to understand the thought process by which the attorneys determined responsiveness. A TAR review will therefore never be completely accurate, and precision will never be 100%.

Low precision, however, is cause for concern about the TAR process used. It may be indicative of the lack of a pattern in the attorneys' review, or perhaps the pattern is not recognizable to the TAR algorithm. It could also be that the pattern TAR identified is overinclusive. A TAR production set with a low precision score is inherently subject to question.

The second metric of a TAR review is the recall, which is the fraction of actually responsive documents in the collection that the TAR review also identifies as responsive.¹⁸ A low recall rate means that the documents produced are highly underinclusive. Recall, like precision, can be ascertained in a number of ways; for example, it can be estimated by attorney review of a sample of documents that TAR identified as non-responsive.

The ideal recall rate should be 100%—that is, TAR successfully sweeps up all the actually responsive documents in the collection. However, as noted earlier, TAR op-

16. Karl Schieneman & Thomas C. Gricks III, *The Implications of Rule 26(g) on the Use of Technology-Assisted Review*, 7 Fed. Cts. L. Rev. 239, 249–50 (2013).

17. If TAR identifies 100,000 documents as responsive and 80,000 of them are actually responsive, then the precision is 80%.

18. If there are 100,000 actually responsive documents in the collection and if TAR identifies 80,000 of them, then the recall rate is 80%.

erates from the hypothesis that responsive documents are in some way similar to each other. The hypothesis may, on the whole, be appropriate for a particular document request, but it will most likely not be true down to every last actually responsive document. TAR is not designed to identify all actually responsive documents, nor should it be expected to.

On the other hand, a low recall rate is indicative of problems with the review, for example, that the pattern TAR identified in the attorneys' seed set review is not appropriately able to identify responsive documents within the collection.

Rule 26(g) provides no guidance on what rate of recall or precision—from 0% to 100%—is sufficient for purposes of a “reasonable inquiry.” On top of that, an acceptable threshold for either recall or precision may depend on the particularities and circumstances of a request. Though a recall or precision of 80% may be appropriate for one particular review, this does not mean that 80% is a benchmark for all other reviews. The recall and precision numbers generated also depend on the method used to calculate them. A recall rate of 60% is itself meaningless without understanding how one arrived at that number; under a different scheme, the recall rate may be higher or lower.

To avoid disputes about how a TAR review is conducted, judges should encourage the parties to discuss and agree ahead of time on the method used to calculate recall and precision, as well as the necessary tolerances. It also may be helpful to the parties and the court for the responding party to supply the relevant statistics when the documents are produced. It is far easier for judges to hold the parties to the values they had previously agreed upon and were comfortable with than for the court to be embroiled in a dispute between parties who had never discussed these issues nor made prior agreements.

Appendix: An Illustrative TAR Order

UNITED STATES DISTRICT COURT

_____ DISTRICT OF _____

_____ DIVISION

Plaintiff,

v.

Defendant.

Case No. _____

Judge _____

ORDER ON USING TECHNOLOGY-ASSISTED REVIEW IN DISCOVERY
OF ELECTRONICALLY STORED INFORMATION IN CIVIL CASES

1. This Order supplements other discovery rules and orders. It streamlines the discovery of electronically stored information (“ESI”) in civil actions, to promote proportionality in that discovery and the “just, inexpensive, and expeditious” resolution of this action. This includes the prompt resolution of discovery disputes, without the court’s involvement if possible and with it as and when needed.

2. This Order may be modified for good cause. The parties must jointly submit any proposed modifications within 30 days after a Fed. R. Civ. P. 16 conference. If the parties disagree over the modifications, they must submit their competing proposals and a summary of the disputes within the 30-day deadline.

3. No later than ____, each party must identify one or more E-discovery liaisons to each other. Each liaison must be knowledgeable about, and will be responsible for, discussing their party’s respective ESI with the other designated liaisons. E-discovery liaisons must be, or have access to, individuals who are knowledgeable about the technical aspects of the discovery of ESI, including the location, nature, accessibility, format, collection, review methodologies, and production in the action.

The E-discovery liaisons must promptly confer with each other and others when needed and must promptly attempt to resolve any disputes without the court’s intervention. If the liaisons cannot resolve the parties’ disputes, the parties must promptly notify this court of the need for its involvement.

4. In responding to an initial Fed. R. Civ. P. 34 request, or earlier if appropriate, the parties must meet and confer about methods to search for and identify ESI that is subject to discovery and to filter out ESI that is not discoverable.

5. Using Technology-Assisted Review (“TAR”) to respond to a Fed. R. Civ. P. 34 production request is considered a reasonable inquiry under Fed. R. Civ. P. 26(g) if

- (a) the scope of the ESI to be searched is sufficiently large; and
- (b) the nature of the request is conducive to using TAR.

6. The scope of the ESI to be searched is considered sufficiently large if, for a given request, the number of ESI documents to be reviewed could not be limited to 10,000.

7. The nature of the request is considered conducive to using TAR if

- (a) the ESI subject to production can reasonably be expected to share relevant similarities;
- (b) the proposed TAR methodology can measure and detect those similarities; and
- (c) the request could not be translated into a set of narrowly tailored search terms.

8. At least ____ days before using TAR to identify the ESI subject to production, the Responding Party must disclose in writing to the Requesting Party its intention to use TAR and the following details about the proposed methodology:

- (a) the name, publisher, version number, and description of the TAR software and TAR coding process; and
- (b) a description of the ESI to be searched using TAR, including
 - (1) the custodians or sources;
 - (2) the ESI data types (such as email, electronic documents, JPEG files, and other types of data);
 - (3) the number of documents to be searched, in total and for each custodian or source; and
 - (4) the responsiveness categories of the ESI produced.

The Responding Party must make its liaisons reasonably available to the other parties' liaisons to address questions about the technical operation of the TAR software.

[Alternative 1]

9. **[Alternative 1A]** The Responding Party must produce

- (a) all nonprivileged ESI that humans reviewed in the process of seeding or training the TAR methodology; and
- (b) the designations the human reviewers assigned to the ESI to show responsiveness to the request.

[Alternative 1B] When the ESI is produced, the Responding Party using TAR must also identify any portion of the ESI production that humans reviewed in the process of seeding or training the TAR methodology.

10. Within ____ days of producing the ESI, the Responding Party must also disclose the statistics showing the validity of the TAR methodology used.

11. Under Fed. R. Evid. 502(d), the production of attorney–client-privileged or work-product-protected ESI does not waive the privilege or protection, either in this action or in any other federal or state proceeding. However, the Responding Party has the duty to conduct a manual review of the ESI that TAR identifies as subject to production, to identify ESI that is covered by attorney–client privilege or work-product protection. The Responding Party must act with reasonable promptness in advising other parties if privileged or protected ESI has been inadvertently produced and therefore cannot be disclosed or used and must be returned.

[Alternative 2]

9. Each party must designate ____ representatives who will work together to seed and train the TAR methodology. These representatives must perform the task at the same time, in the same room, within which the only computing or information recording devices permitted are the devices the experts will use to perform the task. The experts must agree in writing not to disclose the content of nonresponsive or privileged documents they see within the room to anyone other than the Responding Party. They will be subject to sanction for disclosure or improper use.

10. Under Fed. R. Evid. 502(d), the production of attorney–client-privileged or work-product-protected ESI does not waive the privilege or protection, either in this action or in any other federal or state proceeding. However, the Responding Party must act with reasonable promptness in advising other parties if privileged or protected ESI has been inadvertently produced and therefore cannot be disclosed or used and must be returned.

For Further Reference

Paul Burns & Mindy Morton, *Technology-Assisted Review: The Judicial Pioneers* (Sedona Conference 2014).

Elle Byram, *The Collision of the Courts and Predictive Coding: Defining Best Practices and Guidelines in Predictive Coding for Electronic Discovery*, 29 Santa Clara Computer & High Tech. L.J. 675 (2013).

Gordon V. Cormack & Maura R. Grossman, *Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery*, in *Proceedings of SIGIR 2014: The 37th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, 153–62 (2014).

John M. Facciola & Philip J. Favro, *Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled to Work Product Protection*, 8 Fed. Cts. L. Rev. 1 (2015).

Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, 7 Fed. Cts. L. Rev. 1 (2013).

Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and Efficient Than Exhaustive and Manual Review*, 17 Rich. J.L. & Tech. 11 (2011).

Shannon Capone Kirk & David J. Kessler, *The Use of Advanced Technology in Document Review*, in *The Federal Judges' Guide to Discovery* (Electronic Discovery Institute, 2d ed. 2015).

Karl Schieneman & Thomas C. Gricks III, *The Implications of Rule 26(g) on the Use of Technology-Assisted Review*, 7 Fed. Cts. L. Rev. 239 (2013).

The Federal Judicial Center

Board

The Chief Justice of the United States, *Chair*

Magistrate Judge Tim A. Baker, U.S. District Court for the Southern District of Indiana

Judge Curtis L. Collier, U.S. District Court for the Eastern District of Tennessee

Chief Judge Barbara J. Houser, U.S. Bankruptcy Court for the Northern District of Texas

Judge Kent A. Jordan, U.S. Court of Appeals for the Third Circuit

Judge Kimberly J. Mueller, U.S. District Court for the Eastern District of California

Judge George Z. Singal, U.S. District Court for the District of Maine

Judge David S. Tatel, U.S. Court of Appeals for the District of Columbia Circuit

James C. Duff, Director of the Administrative Office of the U.S. Courts

Director

Judge Jeremy D. Fogel

Deputy Director

John S. Cooke

About the Federal Judicial Center

The Federal Judicial Center is the research and education agency of the federal judicial system. It was established by Congress in 1967 (28 U.S.C. §§ 620–629), on the recommendation of the Judicial Conference of the United States.

By statute, the Chief Justice of the United States chairs the Center's Board, which also includes the director of the Administrative Office of the U.S. Courts and seven judges elected by the Judicial Conference.

The organization of the Center reflects its primary statutory mandates. The Education Division plans and produces education and training for judges and court staff, including in-person programs, video programs, publications, curriculum packages for in-district training, and Web-based programs and resources. The Research Division examines and evaluates current and alternative federal court practices and policies. This research assists Judicial Conference committees, who request most Center research, in developing policy recommendations. The Center's research also contributes substantially to its educational programs. The Federal Judicial History Office helps courts and others study and preserve federal judicial history. The International Judicial Relations Office provides information to judicial and legal officials from foreign countries and informs federal judicial personnel of developments in international law and other court systems that may affect their work. Two units of the Director's Office—the Information Technology Office and the Editorial & Information Services Office—support Center missions through technology, editorial and design assistance, and organization and dissemination of Center resources.