A Comprehensive Guide to

# Technology-Assisted Review (TAR)

LIQUID LITIGATION MANAGEMENT, INC.
Logical Solution.
Legendary Service.

## WHAT IS TECHNOLOGY-ASSISTED REVIEW?

Technology-assisted review (TAR), sometimes called computer-assisted review (CAR), is the use of artificial intelligence and algorithms to identify relevant documents based on some element of human training. It is a process that can greatly trim down e-discovery review time because once the computer is trained to recognize tagging patterns, relevant tags are applied to terabytes of documents in a fraction of the time that it would take human reviewers.

The technology behind some TAR platforms, including Liquid Litigation Management, Inc.'s implementation of TAR, uses Latent Semantic Indexing (LSI). This algorithm discerns the contextual content of documents, rather than simply looking for keywords, to draw connections between documents. The TAR process begins when an expert reviewer tags a subset of documents, and based on those decisions, the computer uses machine learning to code the remaining documents.

*TAR is an advanced technology that automates the categorization of legal documents based on iterative human training.*

Although the underlying technology of TAR is being applied in novel ways to the legal industry, technology-assisted review isn't in its infancy. Furthermore, TAR is as much a process and set of best practices as it is a technology, and at this point the industry is still working toward establishing those best practices. In this white paper we will cover effective use cases of TAR, technical explanations, how to analyze and perform quality control, and current best practices based on court standards. Please note that each TAR platform is developed slightly differently and that this paper addresses the topic in the context of our implementation of technology-assisted review as used in Liquid Lit Manager™.

## WHY SHOULD YOU CONSIDER USING TAR AND WHAT DOES THAT PROCESS LOOK LIKE?

Technology-assisted review can be used in various ways, but the main benefit is far and away cost and time savings. TAR is perfect for quickly identifying documents related to a specific issue, such as privilege. It is often used for cases that have large amounts of data and few personnel or time resources to review and produce. TAR helps to cut down on the data volume humans have to review by homing in on responsive documents more quickly. Because typically only one expert reviewer is coding the documents, fewer reviewers are needed. This also ensures a more consistent review as there is only one reviewer's opinion to contend with.

To begin the technology-assisted review process, the human reviewer starts by going through a certain number of documents and tagging them with mutually exclusive tags (e.g. responsive, non-responsive). The human-reviewed data is called the control set. After the control set is reviewed by the user, TAR will enter a second phase to code the remaining document population based on the decisions that the reviewer has made. TAR works on the premise that each document falls somewhere on the spectrum between two opposites. Documents that fall into the middle category are "undecided" and will need further human input. This process repeats, and each iteration incrementally refines the boundaries of what the machine considers to be responsive and applies this to the entire corpus of documents.

In subsequent training iterations, the reviewer will manually review additional documents and compare their designations to the designation determined by TAR. Each confirming or disagreeing decision made provides further refinement to the computer's engine in making future decisions.

When a user is comfortable with the assisted review results, the review can enter its final phase, quality control. The human reviewer can then analyze results and do further rounds of training if necessary. Each time the human codes more documents, the subsequent technology-assisted review becomes more accurate.

*Continued...*

## MOST EFFECTIVE USE CASE SCENARIOS

While technology-assisted review can have impressive results, it is not always the most effective route for all matters. It is important to know what data sets work well with this technology and when to consider other options.

*"The Court recognizes that computer-assisted review is not a magic, Staples-Easy-Button, solution appropriate for all cases. The technology exists and should be used where appropriate, but it is not a case of machine replacing humans: it is the process used and the interaction of man and machine that the courts needs to examine."[2]* — Judge Peck

TAR is **most effective** when:

- The document population is of sufficient volume. For statistic validity, use at least 10,000 documents, and 100,000 documents or more for optimal use.[1]
- The document set predominantly contains text-based documents like emails, letters, and court documents.
- Paired with other ECA and review tools like:
  - › Key-word searching and culling, date filtering and de-duplication
  - › Clustering
  - › Concept-based searching

TAR may be **less effective** when:

- The number of documents is small.
- There are a lot of non-text based documents, or relevancy is based on non-text attributes of documents. For example dates or other metadata fields.

Document relevancy is determined by the textual contents of documents rather than files such as images, blueprints, source code, or other non-text based factors. Therefore, excluding non-text based files is typically advisable. This will make it easier for TAR to predict the reviewer's choices, and some TAR applications will facilitate this practice.

## SELECTING THE CONTROL SET

The first phase of technology-assisted review focuses on creating and reviewing a control set of documents. There are many ways to gather data for a control set, most of which involve sampling:

**Random Sampling** will return a user-specified percentage of the total document population. This is a quick and easy way to get a sample set, but it is not guaranteed to be statistically representative of the population as a whole.

**Simple Random Sampling (SRS)** will return a statistically representative sample based on the total number of documents in the population and chosen statistical parameters such as confidence level and margin of error, which will be discussed later. LLM, Inc. recommends the use of SRS because it offers an unbiased, random selection of documents that is a good representative of the corpus.

**Judgmental Documents** allow you to add documents via a saved search. This method ensures that you add documents to your training set which you know

1 *As the size of a population increases, the sample size using a set confidence level and margin of error will also increase; however, the increase is not linear. The proportion of sample size to population size decreases as the population size increases. For a population of 10,000 documents, a confidence level of 95% and a margin of error of 2%, the required sample size will be about 2,000 documents, or almost a fifth of the whole case which would need to be reviewed by a single reviewer. At 100,000 documents and the same confidence level and margin of error as before, the required sample size increases to almost 2500 documents – only 2.5% of the case instead of 20%. Furthermore, any statistical inferences become stronger as the population size increases because data will become closer to a normal distribution, from which sample size calculations are based.*

2 *Da Silva Moore v. Publicis Groupe, — F.R.D. — (S.D.N.Y. Feb. 24, 2012), available at: www.ediscoverylawalert.com/uploads/file/DaSilva_Moore_11_civ_1279_Opinion_20120224.pdf*

*Continued...*

to be responsive. It includes human judgment in machine learning by pulling from previously coded documents that the expert reviewer believes are good representatives of what documents match one of the binary tags used in TAR.

## RECOMMENDATIONS

- Use Simple Random Sampling to obtain a training set. SRS allows you to draw statistically meaningful inferences about the whole population of documents.
- Use Judgmental Documents to supplement your SRS/ random sample set with known responsive documents.
- Understand how your choice of confidence level and margin of error will affect your results.
  - **Margin of Error** – This is the acceptable difference between the proportion of content-relevant documents in your sample and the proportion of content-relevant documents in the whole population you are willing to accept. A higher margin of error will decrease your sample size, but will give you less information regarding the whole population. We advise setting a default confidence level of 95% with a margin of error of 2.5%.
  - **Confidence level** – This is the percent certainty that the proportion of content-relevant documents in your sample set is the same as the proportion of content-relevant documents in the population, within a margin of error. Increasing the confidence level will increase the sample size, but will also give you better statistical inference into the population. We advise choosing a confidence level between 90% and 99%.

Decreasing the confidence level or increasing the margin of error will make your sample size smaller, but will potentially be a less representative example of the overall population. Also, document populations with a low proportion of responsive documents, or low richness, may require a larger sample or more iterations to properly train the machine. If your data set has a large amount of non-text documents that have been excluded from the TAR process, be aware that you will have a skewed sample of your corpus.

## CODING THE CONTROL SET

TAR looks at a reviewer's coding decisions and bases its own decision-making on that of the coder. In order to ensure cohesive decision-making and to mitigate as much bias as possible, choose one of these methods for tagging:

- **Single Reviewer Method** – One expert team member reviews the control set and any subsequent training phases.
- **Multi-Member Team Method** – A team of knowledgeable reviewers collectively decides how to tag the same document in order to develop a concept at the group level of what constitutes relevancy for the particular case. Coming to a consensus within a group of reviewers is crucial for the efficacy of TAR.

2 *Based on current rulings, set a default confidence level of 95% with a margin of error of 2.5%. This means that, in your sample set of documents, the sample proportion of content-relevant documents is within +- 2.5% of the proportion of content-relevant documents with a 95% certainty. So if you have 20% relevant documents in your sample, then you are 95% sure that the population will consist of between 17.5% and 22.5% relevant documents.*

*Continued...*

In either case, the reviewers should be the team members with the highest level of expertise and knowledge of the case. This is to ensure that coding decisions are both consistent and well-informed. Consistent coding will assist the TAR algorithm to learn faster and more accurately. Furthermore, leveraging the expertise of a knowledgeable reviewer helps ensure that the technology-assisted review choices are based on the most sophisticated decisions.

## FINE TUNING WITH THE TRAINING PHASE

TAR's iterative process compares past results to the current training round in order to improve and refine the decision-making process. There are certain common measurements that indicate TAR's progress towards refinement and are used to further train the engine. Three particularly important sets of document decisions are:

- **Overturn** – Documents for which the human reviewer disagrees with the computer's decision following a training iteration. After each training round, the technology-assisted review algorithm should become more and more accurate, and the overturn, or change in prediction, between each round should steadily decrease.
- **Conflict** – Tagging decisions made by the system that it later changes after further training. After each training round, the technology-assisted review algorithm more closely predicts the human reviewer's coding choices, so the conflict between human and technology-assisted review will decrease.
- **Undecided** – Documents for which TAR was unable to reach a decision. As TAR gathers more information from subsequent rounds of review, there will fewer undecided documents.

Overturn, conflict, and undecided are measures of stability and can be thought of as an overall measure of change between iterations. After every training round, each of these measures should decrease as the coding predictions become more stable. Fewer documents will have their coding decisions switched (conflict), there will be fewer differences between the reviewer and predicted coding (overturn), and the number of documents which cannot be definitively identified will decrease (undecided).

A basic approach of additionally reviewing sets of randomly selected documents would cause small incremental improvements mainly tied to the fact that the user is simply reviewing more documents. Even more powerful algorithms can actively seek out the documents most likely to affect the responsiveness boundary, greatly leveraging the effectiveness of iterative review. The responsiveness boundary pulls in documents that the system has the most questions about, meaning a combination of documents that gets similar scores to both the responsive and non-responsive categories. It may also pull in "conflict" documents for human review. In internal tests, our own Boundary Refinement algorithm increases assisted review's maximum achievable recall rate from the mid 70's to the low 90's, and live usage of TAR with Boundary Refinement has shown similar results.*

## REPORTING, ANALYSIS AND QUALITY CONTROL

It is important to analyze numerical and graphical representations of how effectively TAR learned after multiple iterations: which documents are considered responsive and which are considered non-responsive?

* *In internal tests, after reviewing the control set of documents and completing training, assisted-review documents have an average recall rate of around 75%. After three rounds of using our Boundary Refinement algorithm, the average recall rate is 83%. After five rounds, it is 90%, and after seven rounds, it is 92%. For later rounds, there is only minor incremental improvement. Again, this is echoed by live usage.*

*Continued...*

As part of any advanced review method, Quality Control is an important element to ensure the efficacy and accuracy of your review. Typical quality control measurements, such as having a senior attorney review a statistical sample of documents for accurate tagging, are inherent in the iterative TAR process. The accuracy and quality of the assisted review are statistically quantified, providing the information necessary to proceed with confidence, whether the final results are used to prioritize review or to tag documents directly.

A strong quality control review involves a statistically representative sample using simple random sampling. This includes the all-important step of defining your desired confidence level and margin of error, described on page three of this white paper.

In evaluating how TAR performs on a set of documents, there are several metrics to consider. Here are some metrics to help you understand how accurately technology-assisted review performed and the breakdown of documents in your population:

- **Precision** – the fraction of retrieved documents that is relevant. If technology-assisted review identifies 15,000 documents from a population as content-relevant and of those, 12,500 are actually relevant, then the precision is $\frac{12{,}500}{15{,}000}$ or 83%.
- **Recall** – the fraction of relevant documents that is retrieved. If there are a total of 15,500 content-relevant documents and 15,000 are identified by technology-assisted review, then recall is $\frac{15{,}000}{15{,}500}$ or 97%.
- **F1** – the summary measure combining both recall and precision.[3] According to The Grossman-Cormack Glossary of Technology-Assisted Review, F1 is a "measure of the effectiveness of a search or review effort, which accounts for the tradeoff between Recall and Precision. In order to achieve a high F1 score, a search or review effort must achieve both high Recall and high Precision."[4]

Precision effectively demonstrates how many of the documents TAR finds to be relevant are actually relevant. Recall essentially represents how many of the relevant documents are found versus not found. F1 is a measure expressing the combined values of recall and precision. As with other advanced review efficiency methods, a quality control Percent Change Report will allow you to ensure your quality standards are being met by determining the percentage of documents that have been re-tagged in a second pass review.

## BACKGROUND ON COURT ACCEPTANCE OF TAR

TAR has not been widely used in the past, but a few key court decisions made in the past three years have set precedents as TAR usage becomes increasingly prevalent. This white paper focuses on the most famous cases and explains best practices and guidelines as detailed by the presiding judge.

Perhaps the most well-known TAR case is Da Silva Moore v. Publicis Groupe because it was the first federal case to approve and call for the use of technology-assisted review as a means of identifying relevant electronically-stored information (ESI). On February 24th, 2012 Magistrate Judge Andrew Peck endorsed the use of technology-assisted review and his opinion was later affirmed by District Judge Andrew Carter. In his opinion Peck stated, "[This opinion] does not mean computer-assisted review must be used in all cases, or that the exact ESI protocol approved here will be appropriate in all future cases that utilize computer-assisted review."[5] Judge Peck called for these best practices:

3 *F1 is the harmonic mean, calculated as* $F_1 \frac{2}{\frac{1}{Recall}+\frac{1}{Precision}}$

4 *The Grossman-Cormack Glossary of Technology-Assisted Review (2013 Fed. Cts. L. Rev. 7), available at: http://www.fclr.org/fclr/articles/html/2010/grossman.pdf*

5 *Da Silva Moore v. Publicis Groupe, — F.R.D. — (S.D.N.Y. Feb. 24, 2012), available at: http://www.ediscoverylawalert.com/uploads/file/DaSilva_Moore_11_civ_1279_Opinion_20120224.pdf*

*Continued...*

- The parties agreed to use a 95% confidence level (plus or minus two percent) to create a random sample of the entire email collection to be reviewed for the creation of a control set for TAR training.
- Certain documents were coded using judgmental sampling .
- Review was completed by senior attorneys.
- All coded documents, barring those that were privileged, would be turned over to the Plaintiffs
- Seven iterative rounds were completed and during each, 500 documents gathered by clustering were reviewed to check for new relevant documents identified by TAR.
- After the final round, a random sample of documents coded as non-relevant were reviewed. This determined stability.
- If stability was not reached by the seventh round, additional rounds of training would be required.

In the conclusion of his opinion, Judge Peck clearly stated: "[This opinion] does not mean computer-assisted review must be used in all cases, or that the exact ESI protocol approved here will be appropriate in all future cases that utilize computer-assisted review."

**Re: Actos (Pioglitazone) Products Liability Litigation** is another ground-breaking case in terms of technology-assisted review use in the legal world. District Judge Rebecca Doherty issued a Case Management Order (CMO) for this matter that included a search protocol whose terminology allowed for the use of TAR on the email of four custodians. In her order, Judge Doherty provided specific instructions to both parties for conducting their discovery of Electronically Stored Information (ESI) using TAR. Some of the key guidelines from this order were:

- Both parties will meet and confer to determine the names of custodians from which to pull a randomized sample of 500 documents.
- Common irrelevant documents such as Spam, Commercial e-mail, System files, etc. will be removed from the sample population
- Defendant and Plaintiffs will each nominate experts to work collaboratively to train the TAR software.
- Defendant's experts will review documents for privilege prior to Plaintiffs' experts viewing them. These will be withheld from viewing by Plaintiff's experts or redacted, and a privilege log will be provided. These privileged documents can still be used to train the software.
- Experts from both parties will work collaboratively to determine the relevance of non-privileged and privileged-redacted documents. Defendant's experts will determine the relevance of privileged-withheld documents.
- The initial Control Set of 500 documents uses a 95% confidence level for estimating richness. There is a worse-case error margin of plus or minus 4.3% assuming richness of 50% that would lower the richness of the set. For the software to function effectively, the Control Set must reach the "Statistical" level of validation, meaning it contains at least 70 relevant documents. For a document set with 14% richness and above, a Control Set of 500 documents is sufficient. For lower levels of richness, experts must review additional documents in the Assessment Phase.
- In order to minimize the margin of error on the TAR software's estimates, both parties will continue reviewing documents in the Assessment Phase beyond the "Statistical" level until the Control Set contains at least 385 relevant documents.

Continued...

**Gordon v. Kaleida Health**, a more recent matter, referenced Judge Peck's decision in Da Silva Moore v. Publicis Groupe and highlights the importance of cooperation by both parties. The court was displeased by lack of progress of both parties and suggested the use of technology-assisted review. The Plaintiff was set to begin review with TAR in September of 2012, but the Defendants objected to the Plaintiff's inclusion of their ESI consultants in meet-and-confer for establishing protocol.

Later that month, the Defendants sent their own ESI protocol, to which the Plaintiffs immediately objected. They cited Moore: "'[e]lectronic discovery requires cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI.'" Moore, 287 F.R.D. at 191 (quoting William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 136 (S.D.N.Y. 2009) (Peck, M.J.)).

On October 5, 2012 the Plaintiff filed a motion to compel for a meet-and-confer with their consultants to establish protocol for use of technology-assisted review.

The Defendant finally explained the reason for objecting to a meet-and-confer with the Plaintiff's consultant. It turns out that the Plaintiff's consultant had previously consulted the Defendant. Otherwise they had no qualms with having a meet-and-confer regarding technology-assisted review protocol. The Defendants expressed their awareness and acknowledgement of their obligations and agreed to a meet-and-confer with a different consultant, so Magistrate Judge Leslie Foschio dismissed the Plaintiff's motion to compel on May 21, 2013.

**Other legal matters that reference TAR:**

- EORHB v. HOA Holdings LLC
- Global Aerospace Inc. v. Landow Aviation
- Kleen Prods. LLC v. Packaging Corp. of Am.

## ALTERNATIVE USES FOR TAR

TAR does not have to be used solely for the purpose of identifying documents that will be directly handed over to opposing counsel. The cost and time saving benefits of TAR are equally as great when used for streamlining the review process, especially when paired with other advanced Early Case Assessment (ECA) and review tools.

Clustering the entire corpus of documents and taking a stratified sample from each cluster provides the best possible breadth of examples for the user to make initial decisions on. Consider a small cluster with few documents whose content talks about 'financial ruin'. If each document was equally considered compared to other documents in the population, a 'financial ruin' example document might not always be selected in true SRS. In a stratified sampling, each cluster is sampled, making sure critical content is in front of the reviewer.

Additionally, users can include other technologies and review efficiencies to enhance the control set by selecting to include in it any useful documents found via other means, such as keyword searches, concept searches and linear review. For example, some may prefer to add documents they know to be high-quality responsive or non-responsive exemplars.

TAR can also be used on opposing counsel's production in addition to e-discovery documents. This gives legal teams keen and swift insight into opposing counsel's production at a fraction of the time needed for manual review.

*Continued...*

All of these methods are useful for faster identification of topics, such as privilege, and as a result these methods prioritize review. By finding and reviewing critical content earlier in the litigation process, both law firms and corporations can save time and money and instead focus on their jobs — securing satisfactory legal outcomes.

## CONCLUSION

It is important to remember that each case and its respective documents are unique. Case uniqueness should be considered when using TAR on other cases – it is not inferred that one case's parameters will yield the same results for another case. Case uniqueness can be attributable to many factors, including, but not limited to: volume of documents, types of documents, reviewer(s), parameters chosen, and what is considered relevant.

Follow best practices outlined in this paper as well as guidance given by the courts. Use the technology to benefit your case in more ways than one – whether in review or for quick, initial assessment of opposing counsel's production. However, to paraphrase Judge Peck, TAR is a new tool in the e-discovery toolbox, not a standalone technology or a replacement for people. Most importantly, the courts have taught us that that a party needs to be on the same page as opposing counsel and the court in terms of usage and methods.