



# PERSPECTIVES

---

## **How the Continuous Active Learning (CAL) Process Turbocharges Ranking and Relevancy in Document Review**

Our perspectives feature the viewpoints of our subject matter experts on current topics and emerging trends.

## INTRODUCTION

The eDiscovery industry often employs the benefits of Technology Assisted Review (TAR) and other document analytic methods when addressing the needs of a document discovery effort. The typical application of these innovations ranges from auto-coding documents for specific case issues to concept clustering and near-duplicate identification. When applied in conjunction with consulting support from trusted advisors and the use of industry best practices, these technologies have helped legal teams effectively reduce the cost of document review.

This paper discusses the benefits of using Continuous Active Learning (CAL), which is a more cost-effective, timesaving, and flexible form of TAR. For investigators and attorneys, CAL provides them with an ability to identify the most relevant documents early on in the review process in a manner that helps them build a more effective legal strategy.

## THE TRADITIONAL TAR MODEL

Predictive analytics or TAR has many flavors and application methods. Across all application methods, the utilization of TAR involves the use of statistics to show that the documents within a sample set are representative of the balance of documents that are subject to review. It also demonstrates that certain characteristics of the sample documents can be used to identify similar documents in the remainder of the review set.

Measuring the effectiveness and level of statistical reliability of the TAR model often falls to two descriptive values:

- Recall, or the percentage of relevant documents identified when applying the TAR model, and
- Precision, or the percentage of truly relevant documents vs. false positives identified when applying the TAR model.

The goal of this technique is to limit the number of documents that need to be reviewed by humans and provide a statistical basis that can be vetted and tested for excluding a portion of the documents from the review.

Often, the development of a traditional TAR model can be a time consuming, repetitive, and complex process at a time when there is significant pressure to review the documents as soon as possible in order to meet court-imposed production deadlines, identify the core documents at issue in a given litigation, and ensure that the teams of reviewers selected to review the documents for the case are occupied as soon as they start charging for their time. Until the TAR model reaches satisfactory levels of precision and recall, additional samples, or training sets, are used in an effort to improve the statistical effectiveness of the model. Typically, review will not begin until the documents identified by the model maintain a recall rate that sufficiently meets the needs of the legal team performing the review.

## THE PRIORITIZATION BENEFITS OF CAL

In contrast, CAL is a form of TAR that allows the review to begin immediately with the benefit of keywords, date restrictions, concept search application, and other methods to focus the review on the core issues at hand without waiting for randomly selected sample training sets to be reviewed and tested against the total population of documents. This can save weeks of sample testing and model training on documents that are largely irrelevant to the issues at hand.

CAL works by building a dynamic TAR model that takes into account every decision made by the review team on the documents that the case team believe will hold the most relevant information. The feedback from the ongoing document review is continuously incorporated into the CAL model. Additionally, CAL can be used to continuously prioritize the documents making their way into the review which leads to the gradual increase in the number of relevant and important documents a reviewer sees as they progress through the process. This procedural change is critically important for investigations or matters involving alleged issues that are not fully understood by the legal

review team as it improves the speed that they can access vital case documents in a matter. The change also provides the flexibility to adapt the documents being batched for review to new priorities as new information surfaces.

Brute force document review, or linear document review, has been proven to be less accurate and inefficient when compared to TAR models. Nevertheless, there are still instances when TAR is not utilized by a legal team because of the required upfront (and sometimes significant) time commitment involved in the development of a traditional TAR model. The training effort often needs to be repeated, resulting in a time-consuming process during the most critical days of a case. The time lost through this process can be problematic for attorneys to justify, particularly on investigations or complex litigations when time is of the essence during the crucial early stages of review.

Identifying and acting on key documents that have a significant impact on the legal strategy of a matter adds additional pressure on the review team to speed up the document review. An important document can impact an array of decisions from how to depose certain witnesses to early settlement decisions. Legal teams will often weigh the time and cost associated with the use of TAR models against starting a review more quickly using less sophisticated document culling strategies, like keyword and concept application, date filtering, etc. CAL changes the calculus associated with these decisions, because it provides the benefits of starting the review immediately with the advantages of a TAR model that will limit the total documents to be reviewed on the matter.

CAL turbocharges the speed and effectiveness of a document review. Not only does it prevent the review of multiple training models at the outset of a matter needed to train a traditional TAR model, it also can be used to continuously improve the number of relevant documents a reviewer sees in each batch of documents pulled during a review, which is particularly important during the critical early days of a matter. By continuously improving the precision of each batch a reviewer pulls to review, the model brings to light relevant documents early in the review and provides the legal team with the opportunity to act and make decisions based on the documents identified in the early stages of the document review. Further, dynamically batching documents in this way improves the efficiency of the document review team

by presenting fewer nonresponsive documents within a review batch. Additionally, CAL models are not limited to relevance decisions. They can be created and applied to any document decision being used to classify documents in a review including case issues, compliance notations, hot or importance flags, etc.

## CONCLUSION

Investigations can benefit greatly through the application of CAL models because of their ability to find the most relevant documents early in a review. Likewise, smaller document review efforts can particularly benefit through the use of CAL, because the upfront model training is not required to gain the benefits of a TAR model.

The document review process can be time consuming, expensive, and require a high degree of discerning accuracy. Using TAR models to improve on the costs associated with document review by eliminating large populations of irrelevant documents from having to be reviewed have made them an important component of a successful document review effort. CAL improves the traditional TAR application by eliminating the early sample training process and allowing the review to proceed as quickly as possible. CAL can also be applied to smaller matters to improve the speed and efficiency associated with smaller scale document reviews.

When your next document review kicks off, look for a document review system that offers CAL and dynamic batching of review sets along with a team of trusted advisors to guide you through the process. Not only will your clients enjoy lower costs associated with the document review effort, but the legal analysis may be enhanced by having access to key documents early on in the review.

## ACKNOWLEDGMENTS

We would like to thank Stephen O'Malley for providing insight and expertise that greatly assisted this research.

Stephen O'Malley is a Senior Managing Director and serves as the practice leader for J.S. Held's Digital Investigations & Discovery group within the Global Investigations practice. He has been engaged on some of the largest multinational investigations and has given expert testimony in the areas of analysis and restoration of electronic data, electronic discovery best practices, and testing of related computer software. He is an expert eDiscovery practitioner and data analyst. Stephen has significant experience in major fraud and corruption investigations including FCPA, Ponzi schemes, U.S. Department of Justice and SEC investigations; in multijurisdictional litigations; in provision of evidence for litigation support; and in advanced data analysis.

Contact Stephen O'Malley at [somalley@jsheld.com](mailto:somalley@jsheld.com) or +1 718 510 5617.

This publication is for educational and general information purposes only. It may contain errors and is provided as is. It is not intended as specific advice, legal or otherwise. Opinions and views are not necessarily those of J.S. Held or its affiliates and it should not be presumed that J.S. Held subscribes to any particular method, interpretation or analysis merely because it appears in this publication. We disclaim any representation and/or warranty regarding the accuracy, timeliness, quality, or applicability of any of the contents. You should not act, or fail to act, in reliance on this publication and we disclaim all liability in respect to such actions or failure to act. We assume no responsibility for information contained in this publication and disclaim all liability and damages in respect to such information. This publication is not a substitute for competent legal advice. The content herein may be updated or otherwise modified without notice.