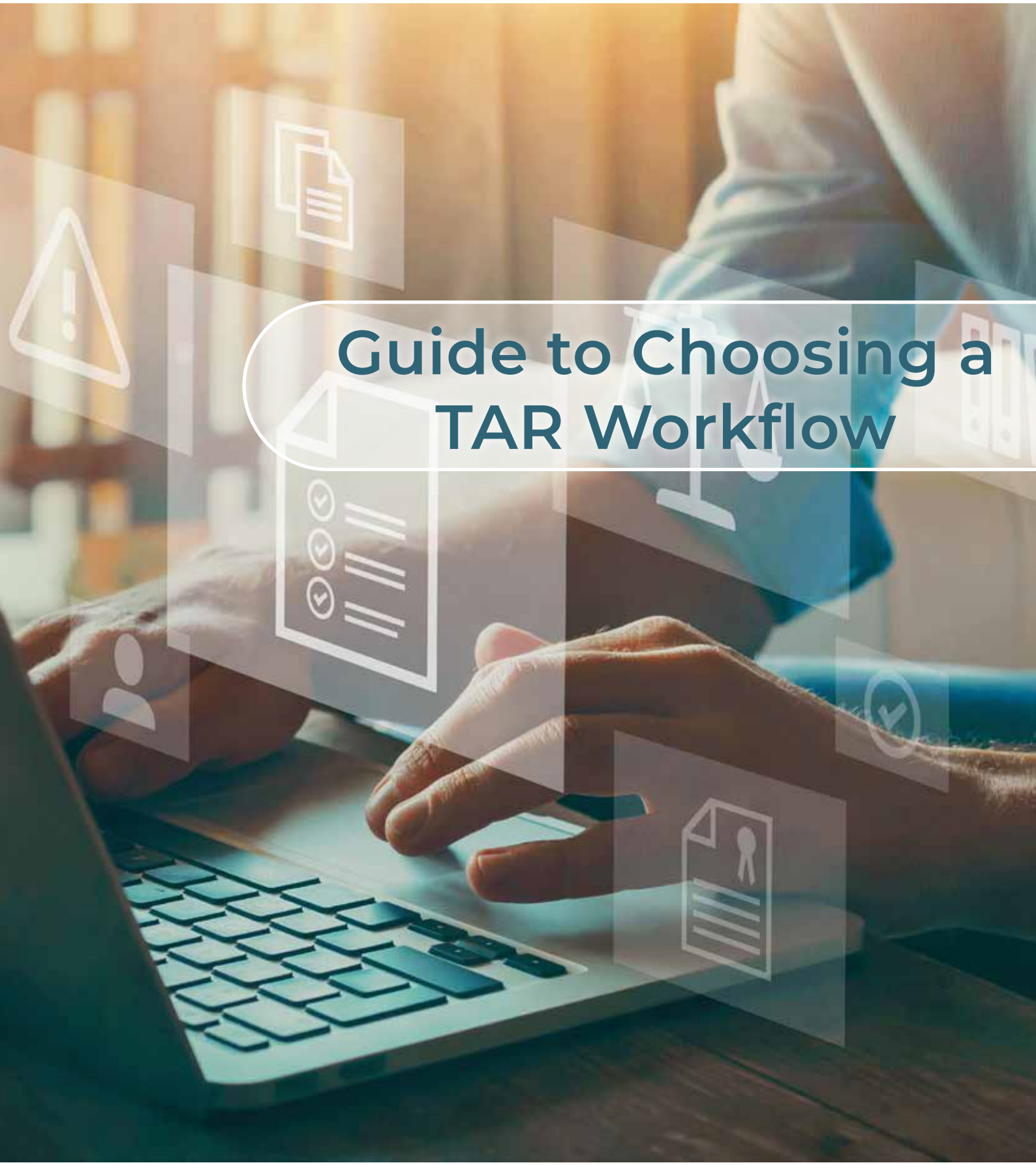
INNOVATIVE DRIVEN



Guide to Choosing a TAR Workflow

Guide to Choosing a TAR Workflow

INTRODUCTION

Lawyers should have a basic understanding of technology assisted review (TAR) so they can speak knowledgeably with clients and providers about whether to use TAR and, if so, which TAR workflow to use. In like manner, counsel must be able to communicate effectively with adversaries and the courts regarding the TAR process their clients have used.

Sifting through available information on TAR workflows and deciphering what information is truly actionable as opposed to a marketing pitch can be overwhelming and time-consuming. This guide assists parties with this process by describing TAR and TAR workflows, as well as factors to consider when selecting the optimal TAR workflow.

WHAT IS TAR?

TAR is a supervised machine learning process whereby human reviewers train software to classify electronically stored documents (ESI) on its own. TAR is an iterative and interactive process between humans and the software. Humans review documents, make a binary decision on how to classify documents, such as responsive or not responsive, and submit the decisions to the software.

The software analyzes the features—usually words, phrases, and metadata—in the document that make it responsive, and then learns from it. With sufficient and appropriate training, the software builds a predictive model that it runs across all documents in the collection and assigns a predictive score to each document. It can then rank documents in order by likelihood of responsiveness.

Guide to Choosing a TAR Workflow

WHAT TAR WORKFLOWS ARE AVAILABLE?

The most common TAR workflows are often known as TAR 1.0 and TAR 2.0. These workflows are also commonly known by other terms. TAR 1.0 is also called "predictive coding" and "sample-based learning." TAR 2.0 is also sometimes called continuous active learning and prioritized review. Some TAR providers also market a TAR 3.0 workflow, while other providers view that as an application of a TAR 2.0 workflow using conceptual sampling for training.

Most producing parties leverage these workflows to defensibly eliminate from review and production documents predicted to be nonresponsive. Producing and receiving parties can also use TAR to prioritize documents for substantive review. Although TAR can be used for other purposes, particularly TAR 2.0, this article focuses on using TAR to search for responsive documents.

TAR 1.0

TAR 1.0 is commonly used in matters where it is appropriate to produce information without an eyes-on review of all documents in the production set. The result of a TAR 1.0 workflow is usually a production of documents that were largely unreviewed by the producing party, allowing the party to comply with production obligations as efficiently as possible. Because most documents are not reviewed, TAR 1.0 uses a sample, or control set, to estimate performance metrics.

TAR 1.0 is most efficient when all, or almost all, documents for the entire project are loaded into the database at the beginning of the project. A subject matter expert (SME) reviews and codes a randomly selected set of documents known as the control set. The control set measures how well the predictive model, described below, is performing throughout the training rounds.

The next step is to train and build the predictive model. To do so, the SME reviews and codes documents in rounds. The documents in the training rounds are usually selected by the software, such as documents it is uncertain how to classify, though human reviewers can also add additional training documents.

After human review is complete for each round of training documents, the predictive model then classifies or ranks all documents in the collection from most to least likely to be responsive. Each time, the software uses the control set to estimate current metrics and improvement from prior training rounds. These training rounds continue until the predictive model has "stabilized," meaning additional training has stopped significantly improving the model.

A cutoff point among the ranked documents is selected, above which documents are predicted responsive and below which documents are predicted not responsive. To determine the cutoff point, counsel often uses estimated recall for responsiveness, at times in combination with other metrics such as estimated precision.

Recall for responsiveness is the percentage of all responsive documents that the TAR software correctly identifies. If there are 100,000 documents in a review that are responsive and the TAR software identifies 80,000, then the recall would be 80%. Precision is the percentage of documents predicted responsive by TAR that are actually responsive. If the TAR software predicts 200,000 responsive documents but only 100,000 are truly responsive, precision would be 50%.

Guide to Choosing a TAR Workflow

Once a cutoff point is established, a review team may wish to do further review of documents predicted responsive before producing them. This review is usually performed only for documents that are potentially privileged or may contain other sensitive information. The documents below the cut-off score generally are not reviewed or produced.

While some parties manually review all documents predicted responsive by a TAR 1.0 review, TAR 2.0 is generally preferred when the goal is to review all responsive documents before producing them.

TAR 2.0

In a TAR 2.0 workflow, training of the predictive model and review of documents is performed simultaneously. Unlike TAR 1.0, which is focused on selecting the most useful set of training documents to efficiently build the predictive model, TAR 2.0 instead is focused on reviewing all responsive documents and avoiding review of nonresponsive documents. Also, unlike TAR 1.0, which relies on a control set to estimate how the model is performing, TAR 2.0 generally measures performance by how well TAR is finding responsive documents, though some samples are still used.

After an initial seed set is coded by human reviewers, TAR selects documents ranked highest as most likely responsive and they are batched for human review. As the review continues, coded documents are continuously submitted to the software as new examples, which further train the model and accordingly refine and improve its predictions of which documents are responsive.

The software continuously re-scores the documents and presents to human reviewers the next- highest scored unreviewed document for review. While the bulk of a TAR 2.0 review is almost always the highest ranked documents, a well-trained TAR model should also include other training documents. New documents can be added at almost any time during a TAR 2.0 review.

As this training and review process continues, the documents selected by TAR for next review round generally become less responsive. The review continues until the percentage of responsive documents becomes low enough that the volume of review to find the remaining responsive documents would be disproportionate under the circumstances. Stopping criteria generally include reaching a low responsiveness rate, desired recall rate, and low elusion rate. Elusion rate is the percentage of unreviewed documents that are responsive which is tested with a validation sample of unreviewed documents.

At the end of the training and review process, all documents identified as responsive will have undergone human review. The unreviewed, low-ranked documents do not undergo human review and are not produced.

SHOULD I USE TAR 1.0 OR TAR 2.0?

There are several factors for a party to consider when deciding whether to use TAR 1.0 or TAR 2.0. Unless a regulatory or government agency mandates a particular TAR workflow and you are unable to negotiate a preferred alternative, there is no bright line rule to apply. Instead, counsel should weigh the importance of the various factors based on individual case strategy, review objective, resources, timeline, and risk tolerance.

Guide to Choosing a TAR Workflow

CONSIDERATIONS FOR TAR 1.0 VS. TAR 2.0

FACTOR TO CONSIDER	TAR 1.0	TAR 2.0
How large is the review universe?	Suitable for about 20,000+ documents	Suitable for about 10,000+ documents
How responsive is your review universe?	Not suitable for highly responsive or highly nonresponsive sets.	Not suitable for highly responsive sets, unless you are using to prioritize review of all documents.
What is the objective of the review?	Compliance	Substantive Development and/or Compliance
Are all documents available at the outset?	Yes (although small additions are ok)	Rolling loads are ok.
Is it an option to produce documents categorized as responsive without reviewing them?	Yes	No
What is the risk of producing unreviewed documents?	High risk	Low risk
What is the risk of producing privileged documents?	Medium to high risk depending on the thoroughness of privilege screens	Low risk
What is the risk of producing sensitive information?	Medium to high risk depending on thoroughness of sensitive information screens	Low risk
Is a separate privilege review needed?	Yes. Review privilege search term hits and families of unreviewed documents categorized as responsive.	No. Simultaneous with responsiveness review.
Can we use first level reviewers?	Prefer coding of control set and training rounds by SME. Can use first level reviewers for any subsequent review of the documents categorized as responsive.	Yes
Will we need to QC responsiveness designations?	Yes	Yes
Are there additional documents that will need review after TAR review completes?	Excluded documents, such as images, media files, very large documents, and uncategorized documents. Documents that hit on potential privilege or other sensitive information.	Excluded documents, such as images, media files, very large documents, and uncategorized documents.

Guide to Choosing a TAR Workflow

ABOUT INNOVATIVE DRIVEN

Founded in 2001, Innovative Driven has built its reputation as the top eDiscovery solutions provider by excelling in complex data challenges. In 2009, Innovative Driven launched ONE as the first eDiscovery software platform to handle Processing, ECA, Review, Production and Case Management in a single environment. This background of innovation has resulted in Innovative Driven developing tools, creative workflow, and a highly secure and scalable infrastructure.

Innovative Driven's mission is to provide the best of breed technology to its clients while delivering solutions with full case lifecycle support guided by expert eDiscovery consulting, world-class processing and hosting services, and managed document review. Innovative Driven provides the most comprehensive solution in the market for large-scale eDiscovery data management and cost reduction.

Innovative Driven is a unique hybrid of a software manufacturer, world-class service provider, and eDiscovery consultancy. This makes Innovative Driven a true information management company and allows it to support its customers across the data lifecycle. Our view is to excel at reducing the cost and risk associated with discovery by shrinking data footprints and creating a more graceful discovery process. Innovative Driven's goal is to help our clients increase efficiency, decrease cost, and gain a competitive edge.

For more information contact us at:

703.533.9200

info@innovativedriven.com

Bloomberg Law first published this article in February 2021.