



Why TAR 2.0

Our Next-Generation Predictive Ranking Engine is Built to Solve Real World Problems and Save on Review Costs

Within a few short years, Technology Assisted Review has revolutionized e-discovery, saving money and time by dramatically (and defensibly) reducing review populations. Even so, first-generation systems had limits that restricted their effectiveness in real-world contexts. Insight Predict is the first of a new generation of predictive ranking engines built for the way e-discovery really works.

TAR 1.0 systems required that a senior attorney do all the training, often having to click through thousands of random documents until the system “stabilized.” Also, these systems assumed you had all of your documents at the start, which is rarely the case. Finally, TAR 1.0 allowed only a one-time training process, with no easy way to continue learning as the review progressed.

The Next Level of TAR

Insight Predict is the next level of TAR. We built it around a “continuous learning” process. As your review progresses, Predict keeps learning and refining its results. It is why our TAR 2.0 engine makes your review more flexible and cost effective. To make it possible, we developed our own ranking database, capable of ranking millions of documents in minutes, rather than hours or days.

Predict employs “Reinforcement Learning,” for which we have a patent application [pending](#). It is an advanced form of continuous active learning, which a respected study proved to be more effective at finding relevant documents than the first generation of TAR [engines](#). The quicker you find relevant documents, the quicker the review is complete and the lower the review cost.

With Insight Predict, learning continues until review is complete. Reviewers, rather than senior attorneys, do the bulk of the training. QC processes flag likely mistakes in judgment. Senior members of the team focus on finding important documents, both to use as training seeds and to help them more quickly understand their case. All the while, the algorithm keeps improving.

Key Differences Between TAR 1.0 and 2.0

TAR 1.0	TAR 2.0
1. One Time Training before assigning documents for review. Does not allow training or learning past the initial training.	1. Continuous Active Learning allows the algorithm to keep improving over the course of review, improving savings and speed.
2. Trains Against Small Reference Set, limiting ability to handle rolling uploads; assumes all documents received before ranking. Stability based on training against reference set.	2. Ranks Every Document Every Time, which allows rolling uploads. Does not use a reference set but rather measures fluctuations across all documents to determine stability.
3. Subject Matter Expert handles all training. Review team judgments not used to further train the system.	3. Review Teams Train as they review, working alongside expert for maximum effectiveness. SME focuses on finding relevant documents and QCing review team judgments.
4. Uses Random Seeds to train the system rather than key documents found by the trial team.	4. Uses Judgment Seeds so that training begins with the most relevant documents, supplementing training with active learning to avoid bias.
5. Doesn't Work Well with low richness/prevalence collections; impractical for smaller cases because of stilted workflow.	5. Works Great in low richness situations; ideal for any size case from small to mega because of flexible workflow.



Ultimately, TAR 2.0 systems such as Predict open the door to a more fluid and flexible approach to TAR, one that works with and adapts to the real world contexts in which e-discovery occurs. Continuous learning means continuous improvement, which translates to quicker reviews and lower review costs.

From the beginning, we argued that continuous ranking/continuous learning is more effective than the TAR 1.0 approach of a one-time cutoff. We have also argued that clicking through thousands of randomly selected seeds is less effective for training than actively finding relevant documents and using them instead. And, lastly, we have issued our own research suggesting strongly that subject matter experts are not necessary for TAR training and can be put to better use finding useful documents for training and doing QC of outlier review team judgments, continuously and on the fly, with the ability to always determine where the outlier pool is shifting as review continues.

The core Catalyst workflow functions by mixing, dynamically and in real time, three streams of information: (1) a relevance feedback stream (aka the continuously updated “best” documents), (2) a contextual diversity stream (aka the “unknown unknowns”, documents that are most about what I know that I know nothing about), and (3) randomly selected documents, to fill any holes in the first two. Furthermore, best practices recommend that if the senior attorney or review manager already has knowledge of certain CSI documents, those be fed to the algorithm to help bootstrap it into a better starting point.

Predict Ranks Every Document Every Time

Most TAR systems start by choosing 500 or so documents to use as a reference set. Training is done against the reference set in the hopes that it will be representative of the larger population. They do this because the hardware would take too long to work against all the documents, which could number in the millions.

Predict’s engine can rank large numbers of documents quickly (e.g., 10 minutes for 700,000 documents). This gives you a more accurate picture of the entire document population and better results when you rank them for review.

Better ranking means fewer documents to review, which saves time and money.