



Cormack-Grossman Study Proves Continuous Active Learning Superior for TAR

*By John Burchfield, DSI
July 16, 2014*

A recent Cormack-Grossman [study](#) – “Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery” – outlined three types of Technology Assisted Review (TAR) methods: continuous active learning (CAL), simple active learning (SAL) and simple passive learning (SPL). The study demonstrated that the CAL protocol achieves higher recall than the SPL and SAL methods.



The CAL protocol involves two interactive tools: a keyword search system and a learning algorithm. Typically, a keyword search is used to create an initial seed set of documents to be reviewed and coded. Once coded, these documents are used to train the learning algorithm, which scores the documents based on likelihood of responsiveness. In iterative fashion, more documents are reviewed and coded until the results can be applied to the entire corpus.

Most predictive coding tools follow either the SAL or SPL protocol. Catalyst Insight Predict is one of the few TAR tools using CAL, which our [friends](#) at Catalyst call TAR 2.0 (making SAL tools – which includes virtually all other TAR technologies – TAR 1.0). TAR 2.0 is a differentiator for us, and this new Cormack-Grossman study demonstrates exactly why it’s a better option.

Since 73 percent of eDiscovery costs go toward review, according to the [RAND Corporation](#), it’s important to use the best TAR methodology possible to reduce these costs. There are two key cost-saving advantages of CAL.

Rolling Data

CAL deals with rolling data more effectively than other methodologies. CAL continuously trains the learning algorithm as more documents are coded. Conversely, passive learning used by most TAR 1.0 systems means the technology is trained one time. So, if you have to add data to the set, you aren’t receiving a more accurate algorithm. In fact, that new data would not be included in the algorithm at all.



The continuous system uses judgmental seeds and relevance feedback, which is better than using random seeds because it requires less human review to achieve the same amount of recall. This especially improves the recall for smaller data collections.

Subject Matter Experts

TAR 1.0 requires subject matter experts to do the seed set coding, which takes time from a senior partner or general counsel. Alternatively, CAL systems can be coded by contract attorneys or lower level members of the legal team because they are just as effective at training the system, according to a [test](#) done by Catalyst. Not only does this save money on billable hours, but it saves time for the senior partners who typically do not want to spend hours looking at seed sets. Using contract attorneys allows the process to get started right away. Then, senior level team members can be reserved to look at any discrepancies from the review team or technology in the quality control set.

At DSi, we have seen success using Catalyst's Insight Predict CAL software. [Click here](#) to read a [case study](#) comparing the costs of this TAR 2.0 software and manual linear review.

To learn more about DSi, visit our [blog](#).
[Original article here](#). © 2015 DSicoverly

DSicoverly®

eDiscovery About People™

About DSi

Serving law firms and corporate legal departments worldwide, DSi is a litigation support services company that provides advanced eDiscovery and digital forensics services. Through five core business processes—DSicollect, DSintake, DSinsight, DSireview, DSisupport—DSi's highly trained staff will help you harness today's most forward technology to gain a competitive advantage. DSi is headquartered in Nashville, Tenn. with offices in Knoxville, Tenn., Cincinnati, Ohio, Charlotte, N.C., Minneapolis, Minn.,

Philadelphia, Penn., Atlanta, Ga. and Washington D.C. For more information, please visit DSi at www.dsicoverly.com or follow us on Twitter at: [@DSicoverly](https://twitter.com/DSicoverly).