Continuous Active Learning (CAL)
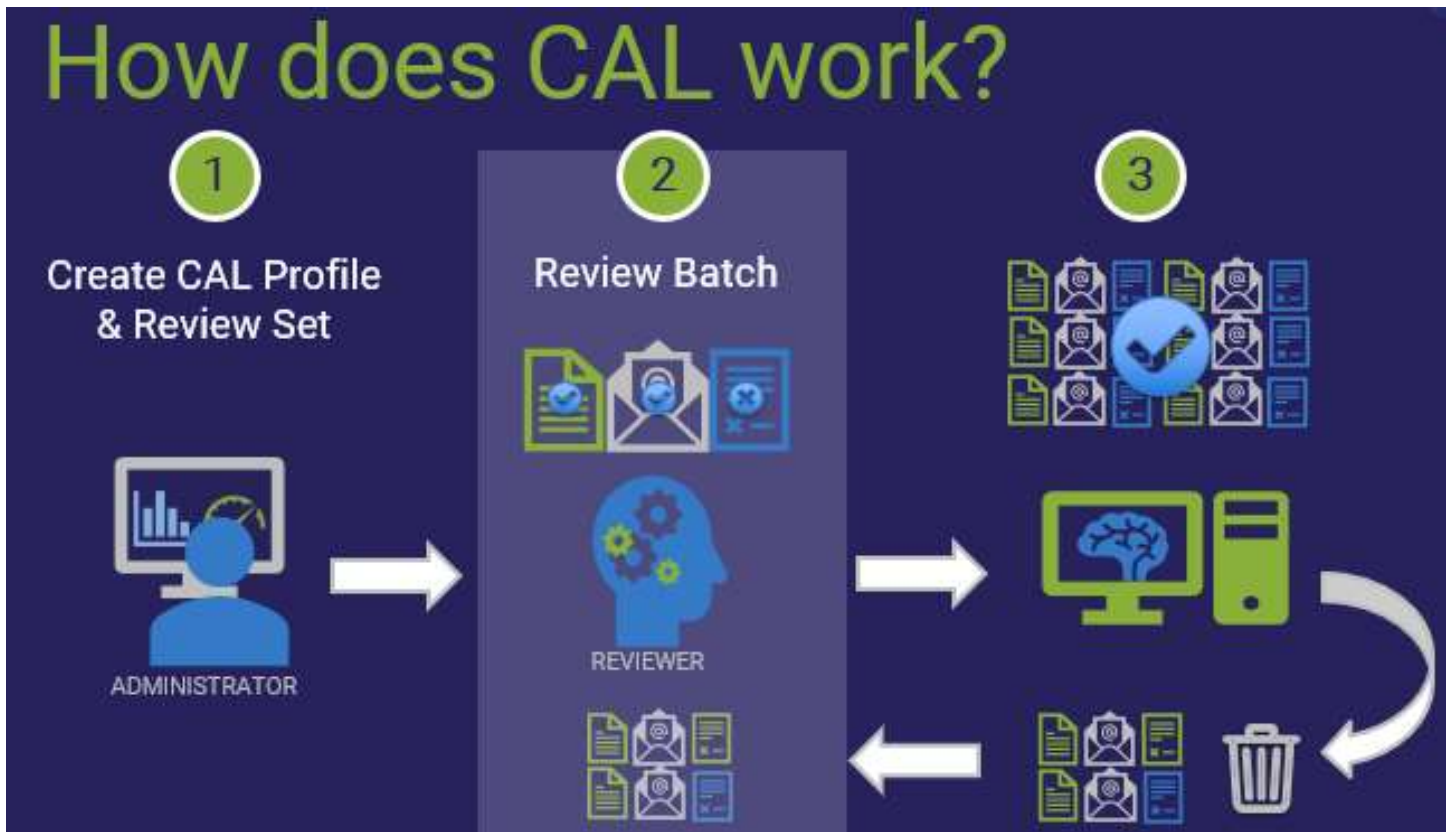
VenioOne

❓ What is CAL?

- A module of the VenioOne platform that leverages **machine learning** to speed up e-discovery review.

- **Training and prediction of relevant documents during the review process** based on the coding done by Reviewers.

- The **direction the industry is heading** in order to efficiently handle larger and larger data sets in cases.

- An alternative to increasing the size of your review teams and a **way to minimize e-discovery costs**.

ONE PROCESS

REVIEW

TRAINING

⚙ How does CAL work?

As with creating any review set, the Administrator establishes review categories (tags) as part of the **CAL Profile**. In addition, the Administrator designates a **primary CAL Category** and sets **Richness and Relevance Thresholds** for CAL to meet.

With CAL, there is no training to be done by SMEs, so Reviewers may **immediately begin reviewing** batches of documents.
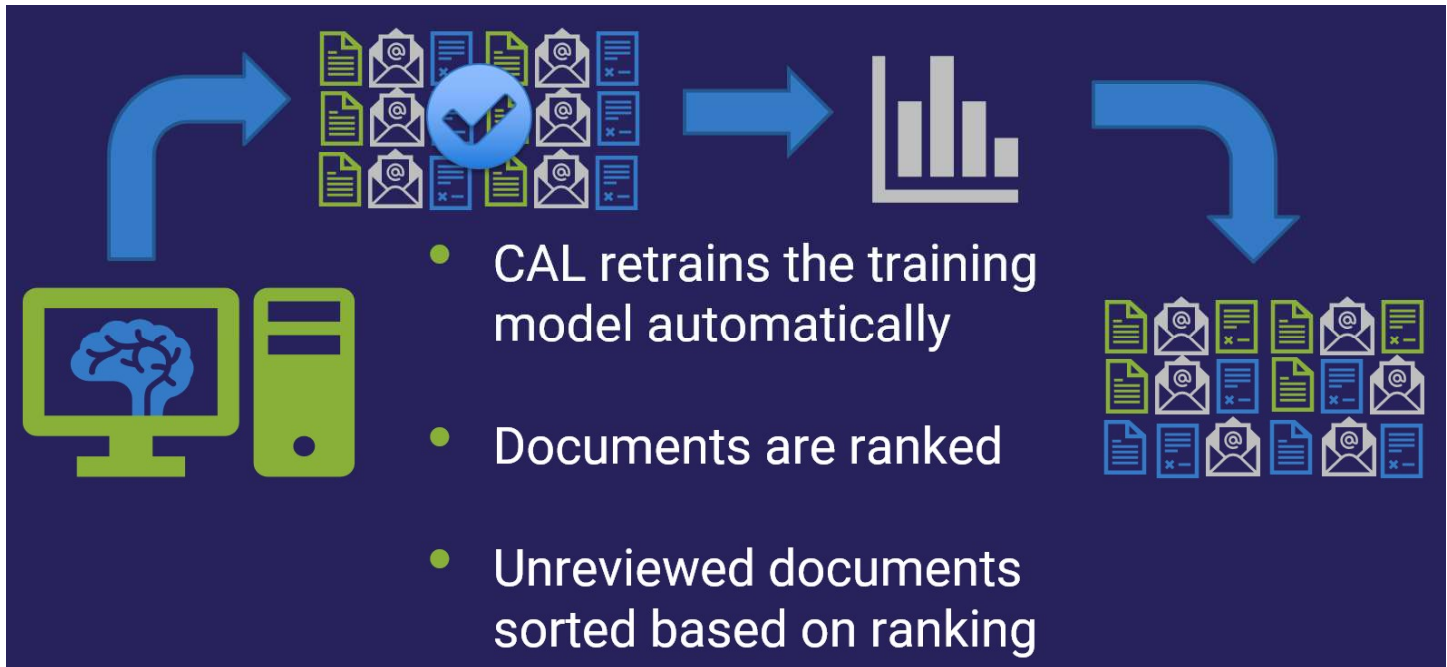
What happens behind the scenes?

As batches are reviewed and checked in, CAL is:

- **Dynamically evaluating the review set**
- **Scoring documents**
- **Recalibrating the model**
- **Re-ranking documents**

This is done in order to present more relevant documents to the Reviewers.

- CAL retrains the training model automatically

- Documents are ranked

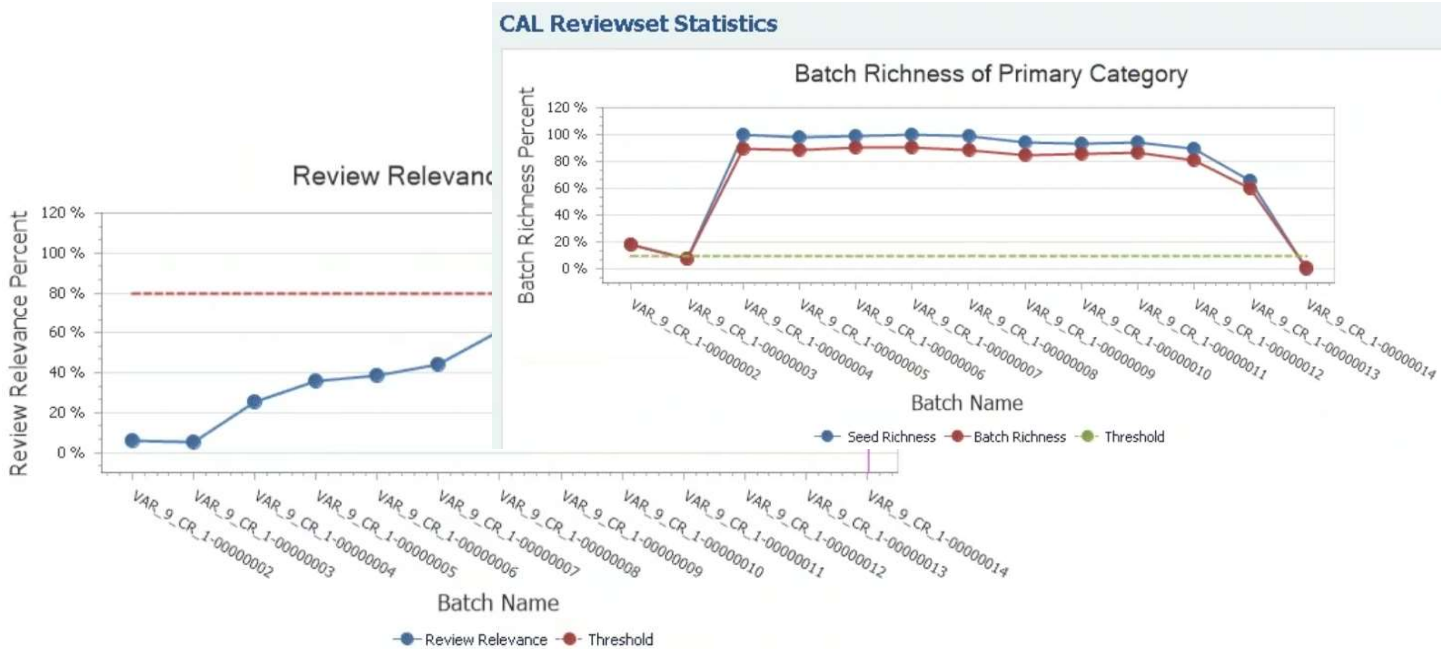- Unreviewed documents sorted based on ranking

Staying on Target

The CAL engine ensures **contextual diversity** in the documents presented as a check and balance.

It may send a mixture of documents, even those with low scores.

CAL Review Progression

The **Batch Richness** and **Review Relevance** statistics below display how CAL learns as batches are reviewed. The number of relevant documents goes up over time but then drops off as the number of relevant documents dwindles down. Eventually, both thresholds set by the administrator are met, so CAL halts the review. Depending on the percentage of relevant documents in the dataset, this can happen with as little as 10% of the dataset being reviewed, so CAL is a huge timesaver!

What other tools are available with CAL?

CAL has a **robust dashboard and reporting** for tracking reviewer progress and CAL's performance.

CAL FAQs

When will CAL be released? Q3 2018

How is CAL licensed? One-time annual fee. No per document or per GB fee.

Which engine is CAL using? CAL uses Venio's proprietary engine and the Support Vector Machine (SVM) algorithm.

What is the real world scenario in which CAL would be the most ideal solution? Any project where you would be paying someone to review documents. CAL is an alternative to growing review teams and a great way to minimize costs. CAL is the direction the industry is going due to increasingly larger data sets. Both sides would need to agree on utilizing the technology.

How is the document set to be used in CAL created? The Venio Review Set Wizard is used just as in a traditional review. For the best results, typically, you would include virtually everything in the dataset, except maybe duplicates.

How does the internal system sampling work? It is like other iterations of assisted review. However, instead of an administrator creating a control set, the system creates it based on the decisions of the reviewers. The system devises a control set based on reviewer tagging.

What is the heart of the inclusion/exclusion process that CAL is using? Is it based on the index? No. CAL creates a model out of the documents based on how the reviewers categorize the documents. It creates the SVM model from that and then scores the remaining documents. The way it sends the documents to the reviewers is based on that document's score, but CAL also sends a mixture, including low score documents, in order to validate and refine the model.

How often does the CAL system re-rank the dataset and does the population size affect that? The re-ranking is done every time a reviewer checks in a batch. Reviewers may continue reviewing as the re-ranking is being done. Batch check-ins do create processing jobs on the back end, so the population size may take additional time for re-ranking.

Does bulk tagging work for sending documents to be categorized? Yes. CAL tags can also be applied outside of the model. Those will be included in the re-ranking when the next batch containing those documents is checked in by a reviewer.

Can CAL be run based on a dynamic folder as the document set? Yes, set the dynamic folder to feed into the review set. However, because dynamic folders are dynamic, they are not ideal for this. Usually, you would use a static set of documents.

Do more documents slow CAL down? Not greatly.

Is CAL approved for use by government agencies? The review tool has been approved – the SVM algorithm.

Are there scenarios where CAL would not be the most appropriate solution? CAL requires the administrator to select one primary issue. Because of this, for identifying multiple concepts or issues at one time, VAR may be a little better.