



epiq security

With the dramatic growth in digital information, corporations and legal teams need new tools to reduce cost and risk.

Predictive coding, a form of technology assisted review (TAR), incorporates well-known technologies to identify relevant and non-relevant documents. An authority on the facts in the matter trains a machine learning system by reviewing a small number of documents. The system then scores the remaining documents, indicating the probability that those documents will be relevant. The review team can then focus their review on the documents most likely to be relevant.

What we offer

Since Epiq is platform-agnostic, we are constantly evaluating the tools and platforms that are available within the marketplace. We acquire or license those best-of-breed technologies that offer compelling advantages to our clients.

An example of this is NexLP Story Engine. We evaluated this product in 2017 and found it to offer advantages that would make a real difference to our clients, including:

- dynamic and interactive data visualisations, so you can see how your data is connected
- sentiment analysis that identifies the tone of a document, not just the content
- summary information about a selected person/ custodian, including who they most spoke to and the topics discussed
- natural language processing to identify key phrases in the document
- advanced filtering tools, including time frames, so you can specifically focus on activities during business, after business hours or late nights
- multiple statistical models, with the best one for your matter automatically selected

Relevant content prioritisation

After loading the documents into the hosted review platform, predictive coding can be used to prioritise highly relevant content for review. This can be achieved

either through review by a subject matter expert or via a team of reviewers briefed by the expert. A typical case only requires training of between 500 and 2,500 documents before the statistical model is ready to use.

Results

The dedicated predictive coding consultants at Epiq guide each project through the process, ensuring proper workflow, documentation, which leads to the best results. While predictive coding is just one piece of a larger solution that is tailored to the specific details of an electronic discovery project, it is one of the best tools for saving time, reducing cost, and improving quality in the discovery effort.

In most cases, the quality and consistency of the result exceeds what a team of experienced reviewers can accomplish if money and time were not a factor.

Is predictive coding right for your case?

In most cases, we recommend the use of predictive coding for projects with at least 20,000 documents to review after standard culling strategies are exhausted. However, there are cases which can benefit from using CAL (Continuous Active Learning) with as few as 5,000 documents when you need to find the most relevant material quickly. Our team of experts can advise you on the most suitable solution for your case.

People. Partnership. Performance.

Contact epiqglobal.com/en-gb/

Simple Active Learning (TAR1.0) workflow

1. Setup: We examine the document population and remove documents inappropriate for textual analysis. The reviewers are trained, and the document-level text is loaded into the TAR tool.

2. Assessment: The review team reviews up to 250 randomly selected documents. This establishes a control set and identifies the richness of the topic in this document population. Richness is the percentage of documents relevant to the topic.

3. Training: The review team reviews between 500 and 2,500 documents. This trains the system to score documents by likelihood of relevance.

4. Batch ranking: The system ranks all documents in the population.

5. Review strategy: The legal team uses an interactive dashboard to make decisions on the level of recall¹ and cut-off scores required for the specific needs of the case.

6. Review: The scores are then used to prioritise or cull the data for review.

7. Quality control and validation: Documents filtered out are randomly sampled and reviewed to validate their low level of relevance.

Continuous Active Learning (TAR 2.0) workflow

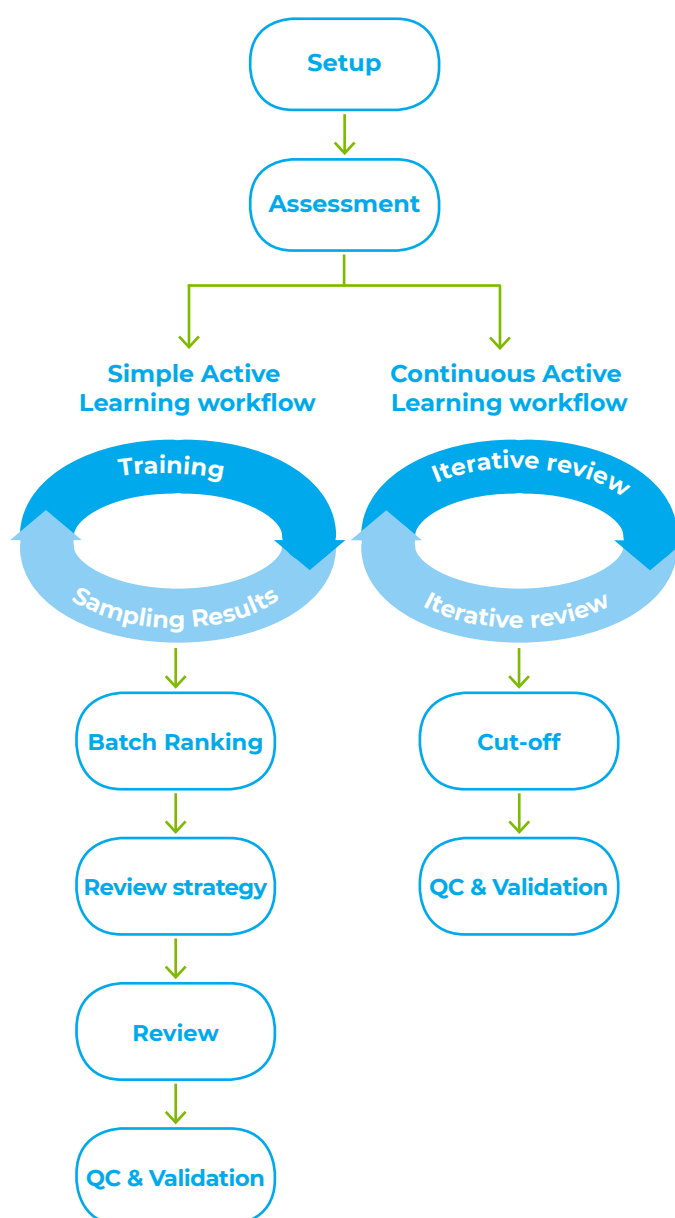
1. Setup: We examine the document population and remove documents inappropriate for textual analysis. The reviewers are trained, and the document-level text is loaded into the TAR tool.

2. Assessment: The review team reviews randomly selected documents until a minimum of four relevant documents have been found. This is initial basis for the statistical model.

3. Iterative review: As the review team reviews the relevant documents, the system continuously updates the statistical model and promotes potentially relevant documents to the front of the review queue.

4. Cut-off: During the review, the percentage of documents ranked as relevant will begin to tail off. We work with your team to select an appropriate point at which to stop document review.

5. Quality control and validation: Documents filtered out are randomly sampled and reviewed to validate their low level of relevance.



¹ Recall is the percentage of the total predicted relevant documents that will be returned by the selected criteria.