AIDA Case Study

Document Review of 2.4 million Documents with AIDA's Advanced Machine Learning Technology – Faster & Cheaper with Better Precision and Recall

The Problem

In the early phase of discovery, data often changes as search terms are negotiated or document requests being updated. Traditional TAR software requires significant human labeled data to retrain itself when that data changes. In order to accelerate review and adapt to data changes, Sullivan & Cromwell used AIDA to review an initial collection of 2.4 million documents for a class action lawsuit. AIDA allowed the law firm to flexibly and continuously update the prediction model without any additional human labeling as search terms and document requests were being negotiated during the early stage of review.

The Solution

AIDA's advanced AI engine largely learns from unlabeled case data and natural language instructions for responsiveness and privilege. With limited training examples, AIDA can achieve high accuracy for both responsiveness and privilege document review. AIDA's AI model can effectively adapt to data changes with little to no human labeling effort. Each time the opposing party requests search term updates or new custodian data, AIDA is able to provide prediction updates within one day. As a result, the document review process is more predictable and more efficient.

The Results

AIDA achieved 93% Precision and 82% Recall in detecting responsive documents and 70% Precision and 95% Recall in detecting privileged documents with only 400 hours of human review (10 reviewers in 5 days). After that, every time the search terms were modified, the model updated its prediction within a day without additional human labeling. The AIDA outputs were imported into Relativity for 2L review and production. The case team not only grasped the key facts of the case early on, but also saved at least one month of review time, and hundreds of thousands of dollars in review cost.

Recall: percentage of responsive (privileged) documents that are correctly identified as such

Precision: percentage of predicted responsive (privileged) documents that are correctly identified as such.