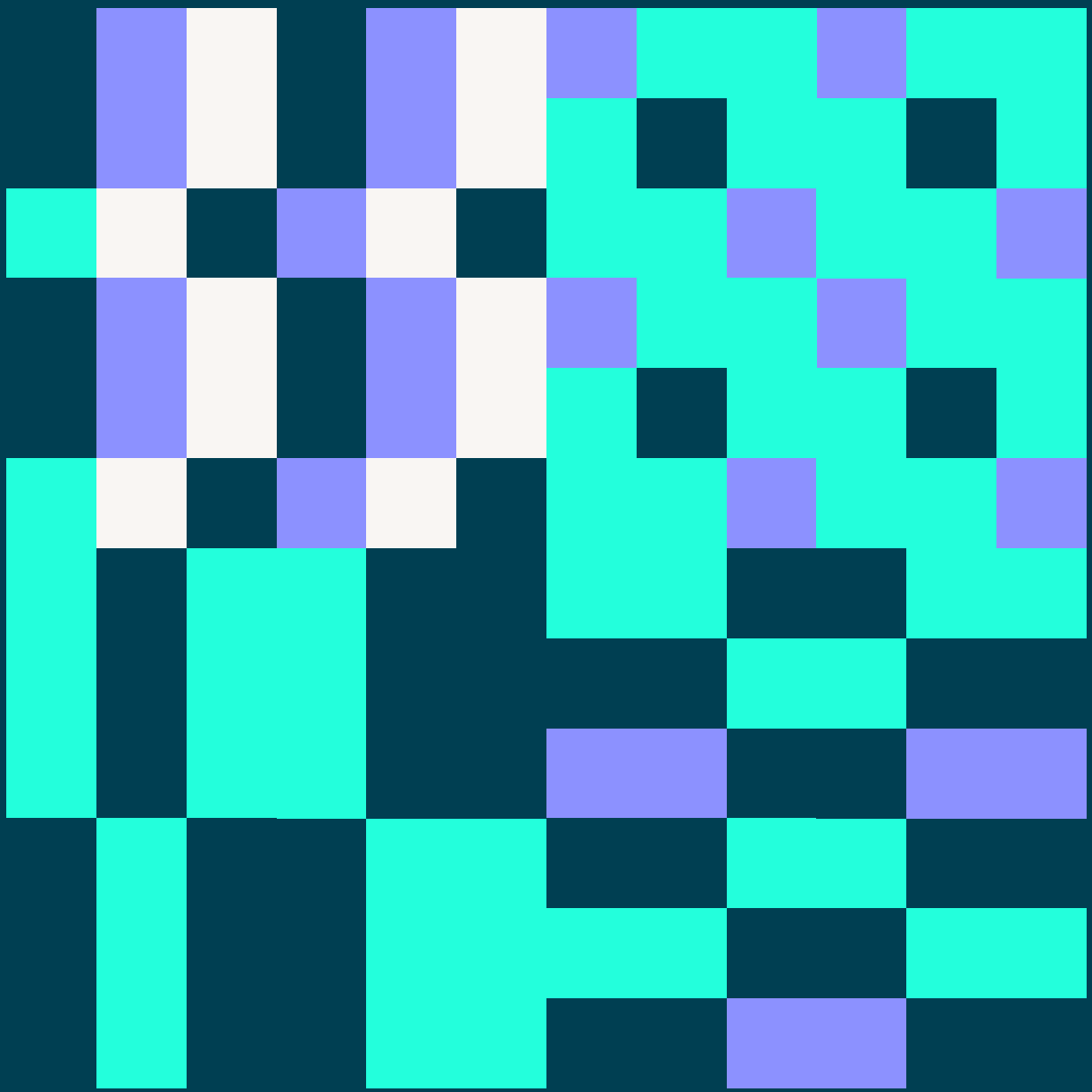
SKY DISCOVERY ■ RESOURCE

# Leveraging Continuous Active Learning



# In this guide we explore how lawyers can leverage Continuous Active Learning for their document review process.


| | |
|---|---|
| Introduction | 3 |
| Documents best suited for CAL | 4 |
| How CAL works | 5 |
| The elusion test in Relativity | 6 |
| The Courts and Regulators position on CAL | 7 |
| How CAL should be considered in practice | 8 |
| Key benefits of CAL | 9 |
| Helpful resources | 10 |
| Solutions and Impact | 11 |
| How we help | 12 |

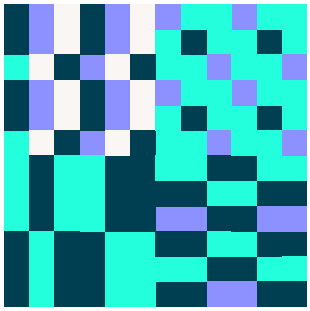



| Reading time | 10 minutes | Page count | 12 pages | Word count | 1,302 words |
|---|---|---|---|---|---|

Nowadays, nearly everything is discoverable. Innovation remains a priority for lawyers, but collecting and organising data is often complex and inefficient. Conventional document review methods are increasingly less viable due to time and cost demands.

A significant breakthrough in addressing this issue is Continuous Active Learning (CAL), a machine learning-based process that transforms how lawyers approach document reviews.

CAL within a platform such as Relativity utilises active learning principles by training a model to prioritise the most informative documents for reviewers. This method enhances the efficiency and accuracy of the review process, ensuring that relevant documents are highlighted. As human reviewers interact with the system, the CAL model continuously learns and refines its predictions based on their most recent inputs and feedback.



# Documents best suited for Continuous Active Learning

### Well-structured documents

*Formatted text*

Documents with clear, well-structured formats, such as contracts, legal filings, and formal correspondence, are ideal for CAL. These documents usually contain distinct sections with headers that can help the CAL model easily recognise relevant patterns and keywords.

*Consistent layouts*

Documents with consistent or uniform layouts, like forms or standard contracts, enable the CAL model to learn and predict relevance more efficiently since the information is consistently presented across similar documents.

### Text-rich documents

*Substantial text content*

Documents with a large amount of text offer more data for the CAL model to analyse and learn from. This includes long emails, detailed reports, comprehensive memos, and other text-heavy files.

*Detailed descriptions*

Documents containing thorough descriptions or extensive discussions of topics relevant to the case are particularly useful for training the CAL model to recognise relevant content.

### Unsuitable documents

Documents with low-text content are not suitable for a CAL review. Documents with minimal text, such as images, drawings, or scanned documents that have not undergone Optical Character Recognition (OCR), are less ideal for CAL. Without searchable text, the model has limited information to analyse.

# How Continuous Active Learning works

### Initial Review

At the beginning of a review project, CAL trains its model using all relevant documents available. This typically begins with tagging documents as 'Relevant-Yes' or 'Relevant-No'. Only five documents in each relevance category are required to initiate the training process.

### Model Training

The model trains on a subset of tagged documents to understand relevance and identify characteristics that make documents relevant to specific queries. Concentrating on these tagged documents improves the model's ability to assess and rank documents.

### Creating a queue of documents

The model identifies documents that are predicted to be relevant for further review and analysis by users or evaluators involved in the process. The queue aims to improve document handling efficiency by highlighting only the most important information for consideration.

### Human Involvement

Reviewers examine these documents and tag them as relevant or irrelevant, providing real-time feedback to the model. They use the four corners rule, assigning relevance based solely on the document's content and not considering metadata or other external factors.

### Model Update

CAL continuously updates its understanding and refines predictions with each new batch of feedback. This process enhances accuracy and effectiveness, adapting to better meet user needs. This iteration cycle can repeat throughout the entire eDiscovery process.

### Validation

As the review queue of documents accumulates irrelevant items the model can be validated using an elusion test. This quality control measure assesses the effectiveness of a review and identifies what proportion of relevant documents may still be undiscovered in the unreviewed set.

# The elusion test is integral to ensuring the thoroughness and defensibility of the document review process.

### The elusion test in Relativity

This test provides a statistical measure of how many relevant documents might still be hidden in the unreviewed data, offering a safeguard against prematurely ending a review and missing critical information. This is how the elusion test typically works within the Relativity workspace.

### Selection of a random sample

Once a substantial portion of the document set has been reviewed using CAL, an elusion test begins by selecting a random sample of documents from the remaining unreviewed set. The size of this sample may vary depending on the total volume of documents and the desired confidence level of the test results.

### Manual review of the sample

The legal team conducts a thorough manual review of the sample to determine if any relevant documents exist. The main goal of this process is to find any important documents that the CAL model may have overlooked or failed to detect during its automated assessment.

### Analysis of results

After the manual review, the results are analysed to calculate the elusion rate, which is the proportion of relevant documents found in the sample compared to the total number of documents in that sample. This rate helps estimate the prevalence of relevant documents that remain in the unreviewed set.

### Decision-making

If the elusion rate is low, it suggests that the CAL model performs effectively and that most relevant documents have likely been identified and reviewed. In this case, the review team may decide to conclude the review process. If the elusion rate is high, it indicates that potentially more relevant documents are still unreviewed.

# The Courts and Regulators position on CAL

**Courts**

Although Australian case law is dated, it acknowledges and approves the use of machine learning technology in these matters:

*McConnell Dowell Constructors (Aust) Pty Ltd v Santam Ltd & Ors, Vickery J*
*Money Max Int Pty Ltd v QBE Insurance Group Ltd*

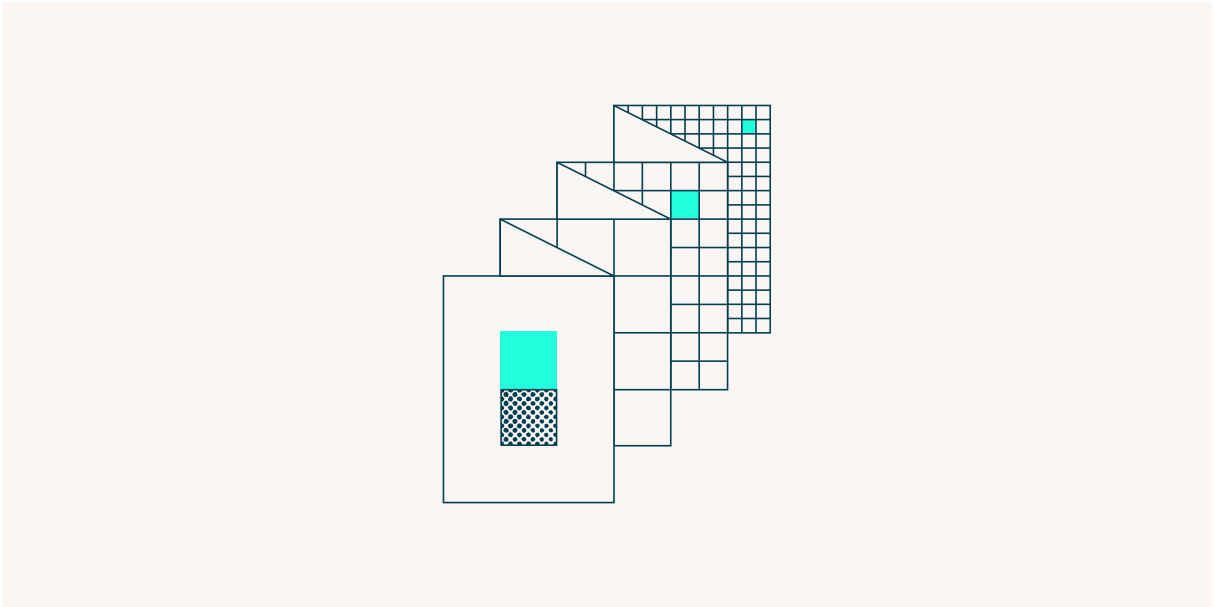
It states, "...the very large number of documents involved in the proceeding calls for special management" and that "... traditional manual discovery of the Plaintiff's documents is not likely to be either cost-effective or proportionate."

Practice Note SC Gen 5 and the Federal Court's GPN Tech recognize "Technology Assisted Review", also known as "Continuous Active Learning", as an approved method for searching and reviewing documents.

**Regulators**

Regulators have already adopted advanced searching technology and used CAL for some time. As per ASIC's Information Sheet 242 (INFO 242). "There are different approaches to reviewing books and identifying the documents that fall under the notice to produce". For example, you may undertake:

a) a manual review;
b) keyword or concept searches; or
c) a technology-assisted review, such as predictive coding



## How Continuous Active Learning should be considered in practice

Document exchange protocols should include a clause allowing reviews to be conducted using CAL. To prevent lengthy disputes regarding a CAL review, it is beneficial to agree on a basic CAL protocol with other parties. The CAL protocol should be flexible and permit parties to use different platforms for conducting a CAL review.

A CAL model designed for making discoveries can also be used to review incoming discoveries from other parties, helping to prioritise the most relevant documents. Australian courts appear reluctant to intervene in disputes about CAL, leaving it up to the parties to determine if it is the best review method while seeking guidance from their eDiscovery consultant.

# Key benefits of Continuous Active Learning

### Efficiency

Continuous Active Learning can greatly reduce the time and effort needed for document review by prioritizing the most relevant documents first.

### Accuracy

CAL increases the likelihood of discovering relevant documents early in the review process, accelerating the workflow and significantly improving the overall quality of the review conducted.

### Cost-effectiveness

By significantly reducing the volume of documents that require manual review, CAL can lower the overall costs associated with the eDiscovery process. This not only streamlines the workflow but also enhances efficiency in legal proceedings.

# Helpful resources

As specialists we continually invest in R&D and best practice so we can advise our partners with confidence. These insights culminate in helpful resources and references for lawyers and decision-makers.

### Data Identification Questionnaire

Our questionnaire aims to help you quickly and accurately identify data potentially relevant to your matter. The information captured from key stakeholders will facilitate the development of a collection plan and enable its swift and defensible execution.

Learn what to consider

### Draft Exchange Protocol (Australia)

This reference is used by our teams on most disputes in most jurisdictions within Australia. The template provides a starting point for developing a protocol that governs the exchange of documents for Australian disputes.

Learn what to consider

### Practice direction by jurisdiction (Australia & UK)

Reference our index of all Australian and UK eDiscovery practice directions.

Learn what to consider

### Learn more

| | |
|---|---|
| Negotiating a document exchange protocol with opposing party | ↗ |
| Reducing discovery obligations with another party | ↗ |
| Developing an appropriate review workflow | ↗ |

# AI or otherwise, when new challenges arise, we find practical, accurate and defensible solutions.

## Our growing AI capabilities

**Chat**

Summarise, translate and label documents using natural language prompts.

**Mass Action**

Prompt, record and reuse multiple document review queries simultaneously.

**Validate**

Review, summarise and fact check document references to supporting evidence.

**Compare**

Review and summarise document similarities or differences in a structured, searchable format.

**Chronology**

Organise, link, track and review an automated sequence of events from documents.

**Scan**

Recognise and tag objects in images and convert them to structured, searchable data.

**Extract**

Capture and populate data from templated forms into structured, searchable data.

**Transcribe**

Extract audio and video files and organise them into searchable transcribed data.

**Translate**

Translate and maintain context in documents in their original format in up to 100 languages.

**Review (aiR)**

Locate material related to legal issues important to your case strategy.

Leverage AI on your next matter. Our solutions ›

## Solutions and impact

### Extracting data from forms using AI workflows

Investigation, Process, Enhance, Review, Sky Solution

Read

### Leveraging continuous active learning in large scale document review

Dispute, Analyse, Review, Sky Solution

Read

### Migrating an active eDiscovery project from another provider

Dispute, Regulatory, Investigation, Process, Sky Solution

Read

Client success ↗

# You need a team with a balance of legal, eDiscovery and technology expertise, this is who we are.

Our expert team of lawyers and technologists are available to assist you with navigating all stages of your matter, from the first meeting, through scoping, to completion. We focus on technical solutions so you can focus on the law. Find out how we help.