AVANSIC®
eDiscovery + Digital Forensics

Artificial Intelligence:
Not As Scary As You Think

Introduction

Artificial Intelligence (AI): it evokes robots, coding, zeros and ones, and seems like it belongs in the domain of computer scientists. In reality it is just computers performing tasks that previously required human intelligence and in fact, you're probably already using elements of AI both in your legal practice and everyday life. The experience you have through a platform like Amazon is different from the person sitting next to you at the office; this is an example of AI at work. An eDiscovery example of AI is mining a database to find a document similar to one you're already looking at.

Advanced Analytics: AI in eDiscovery

AI in Small, Medium and Large Cases

Why it isn't so Scary

Advanced Analytics: AI in eDiscovery

In the eDiscovery and forensics world, AI is often referred to as advanced analytics. Familiar terms that reference AI in the legal and eDiscovery world include analytics, clustering, predictive coding, TAR, CAL, find similar, email threading, near-dupe, data mining, and data aggregation. These techniques can be performed with simple tools that range from Excel to specially designed analytics engines for document review platforms. At its most basic, advanced analytics creates charts and graphs of data to understand what lies within. At its most complex, AI is software making decisions about documents without human interaction.

In modern review tools, all cases have some form of advanced analytics. It is common for all data sets to be de-duped, near-duped, email threaded, and similarity calculations performed as a part of initial processing. This is done on data sets of all sizes and typically does not incur additional fees.

AI in Small Cases

For small cases, these built-in operations previously mentioned are generally enough to assist in organization and speedy review of the documents. For instance, there is no need to look at a document that is nearly the same – at 99 or 100% similarity - twice. It can be coded once and the coding can be applied to the near dupes.

However, context-sensitive coding (such as privileged review) will still require human consideration or searching. For an example of context-sensitivity, if you review a PDF file that was in the My Documents folder and determine it is not privileged, but that same PDF appears in an email attached to outside counsel, it may be privileged.

AI in Medium Cases

In medium-sized cases, apply all the AI techniques from small case still apply. In addition, using charting and graphing of email subjects and email timelines can help to narrow focus for review. For instance, in looking at a pie chart of email subjects sorted by the largest number, there might be subscriptions to journals, news sources, and other alerts irrelevant to the case and therefore unnecessary to review. In another example, a pie chart graphic of email From domains, might show a significant number of emails from shopping sites, which can most likely be removed from the relevant set.

Once clearly irrelevant material is removed in bulk, the application of document clustering will reveals trends in document content. Clusters of relevant documents can be selected and sampled to determine if it warrants human review. In this medium case example, additional cost was only a few human hours. By using most of what is available in advanced analytic techniques, an efficient review set is created that will likely provide responsive data for human review.

AI in Large Cases

For large cases, artificial intelligence can dramatically reduce review team time and improve results. If the above workflow in the medium case example still provided too many documents to economically review within the required timeframe, predictive review, predictive coding, TAR, or CAL techniques are the next layer of advanced analytics.

Regardless of which technology is chosen, humans are training a computer to build an algorithm that will be applied to documents that have not been reviewed by humans, allowing the computer to apply document-level coding. The differences between the techniques are how training, quality control, probabilities are applied.

AVANSIC®
eDiscovery + Digital Forensics

For instance, a person most knowledgeable about the case uses these techniques to begin coding documents. The computer is observing and learning from each document that individual codes until the computer determines that its precision and recall are high enough to continue unassisted. Note that there is risk that the knowledge about the case changes during the course of review in a way that would require re-training the computer. Understanding this is paramount to the success of these predictive techniques.

Why It Isn't So Scary

AI adoption has been slow in the legal community for a variety of very understandable reasons: that it's too expensive (especially for smaller cases), it's hard to understand the technology, it's hard to trust results that can't necessarily be directly tested, and the thought that it will replace billable hours.

Although it has taken some time, the court system is starting to see the benefit for using AI in eDiscovery, specifically to accept that the principles of AI should apply especially in large matters. There are also an increasing number of firms that use this technology to great advantage which creates a larger pool of people that both use and understand the software that are not statisticians and computer scientists. From a tool-use perspective, AI has become easier to use through wizards and default settings that have been wrapped around legal technology software packages.

Cost can be impactful in some cases, but the easiest portions of AI utilization are included in most software packages (see above in the small and medium case sections). Many legal technology software packages include some elements of artificial intelligence and some of those have very reasonably priced licenses. Although there may not be as many attorney hours spent in document review, those that are can be much more fruitful, freeing additional time to be spent on writing, litigating, and other skills that attorneys excel in.

Finally, knowledgeable vendors can be very helpful in determining the type of AI that can be used in certain cases, especially if there are time and cost restrictions. Using these resources can greatly impact the success of an eDiscovery project.