REDGRAVE DATA

DATA. DONE DIFFERENTLY.

Technology Assisted Review (TAR) leverages machine learning and artificial intelligence to more efficiently sort through large volumes of digital documents to identify those relevant to a specific legal case.

What is TAR?

- While there is a lot of technology used in document review, the term technology-assisted review today almost always means the use of supervised learning, a form of machine learning and thus part of the field of artificial intelligence (AI).
- "Supervised" means that the system is trained by example. One or more attorneys review a set of documents and code them (for example as Responsive vs. Not Responsive) as examples for the software. These are called the training documents.
- The supervised learning algorithm analyzes the coded training data and produces a statistical model (called a classifier) that predicts coding decisions that attorneys would make on new documents. It can also estimate its own confidence in its predictions, and sort documents based on that confidence. If those predictions are accurate enough, that reduces or eliminates the need for the attorneys to review more documents.
- In a further twist, the machine learning software can find the training documents from which it will learn the most and suggest those to the attorneys for review. This is called active learning and further reduces the number of documents to be reviewed.

Process Over Software

- While TAR is about technology, the success of a TAR effort also depends on the larger process in which TAR is embedded.
- The foundation of an effective eDiscovery strategy begins with the negotiations by case attorneys. The negotiated custodians, date ranges, file types, and keyword filters (if used) affect both the volume of data, and how much of its review can be aided by TAR. Decisions about requests for production affect the difficulty of the classification task TAR is faced with. The setting of deadlines strongly affects the practicality of different workflows.
- After the data collection is defined, the focus shifts to the overall review structure, including choice of which documents will have TAR applied, what aspects of review (e.g., responsiveness, privilege, issues; first level review vs. quality control) will use TAR, composition of review team, batch size, and so on.
- A final crucial choice is to specify the TAR workflow, which controls how manual review and machine predictions interact. There are two main classes of TAR workflow:

- ➢ TAR 1.0 workflows have two phases. In the first phase, active learning iteratively selects documents for their informativeness for training purposes, and reviewers code those documents. The resulting coded training set is used to produce a classifier. In the second phase, the classifier is used to select a relatively large set of documents. That set of documents might then get reviewed (typically by lower cost reviewers than those used in training), or just a privilege screen, or might be produced without review.

- ➢ TAR 2.0 workflows omit the second phase. Instead, on each iteration a new classifier is produced and used to select the documents most likely to belong to the category of interest. Those documents are coded and serve both as the output of the TAR review and as training data for the next iteration. The process is continued until sufficient documents of interest have been found.

- There are also hybrid approaches that mix characteristics of both TAR 1.0 and Tar 2.0.

Evaluation is Paramount

Evaluation is key in both managing the TAR process and certifying how well it worked.

- A key evaluation method is random sampling, which involves selecting a subset of documents from the entire dataset at random, coding them, and using them to estimate the effectiveness of the TAR process. Such estimates may be used in deciding when to end a TAR workflow, and demonstrating to other parties that the resulting coding of the documents has met established objectives (e.g., for proportion of relevant documents found). However, it's just one of many methods used to understand how well both the technical and manual parts of the TAR process are performing.

- Evaluation not only plays a crucial role in managing individual projects but also in refining and enhancing technology for use in future projects.

The Future of TAR

- Generative artificial intelligence (GenAI) is a major new development in AI. GenAI systems can produce, in response to natural language instructions or "prompts", fluent natural language outputs.

- GenAI is being explored as a way to improve conventional TAR workflows based on classification, as discussed above. Benefits of GenAI in this role are the reduction or elimination of the need to label training examples, and the ability of GenAI systems to point out relevant passages and provide explanations of its classification decisions.

- However, GenAI also provides capabilities well beyond classification. GenAI systems can provide summaries of long documents and, to an increasing extent, sets of documents. GenAI's ability to synthesize fluent language is being explored for privilege log production. GenAI is also being used behind the scenes to improve language processing tools for name extraction, redaction, and related tasks.