



BR SPOTLIGHT

Questions to Ask Legal Tech TAR/PC Speakers and Vendors

OK, so you're going to Legal Tech in New York City and you want to learn as much as you can while you're there. Here are some questions you can ask the speakers and vendors who are promoting the use of Technology-Assisted Review ("TAR") or as it is sometimes called, PC or Predictive Coding. The objective with these questions is to probe some limitations that they may not otherwise want to discuss:

- **Text Dependency for Classification.** *"How does your system analyze document files that don't have associated text?"* Note: don't be misled by statements about using metadata in lieu of the text that is apparent on the face of the documents. Metadata is often incorrect and rarely provides a meaningful way to differentiate among other objects with similar metadata.
- **Text Dependency for Collection.** *"Under your system, how are documents initially collected, do you use key word searches?"* If Yes, then ask the follow-up question: *"Doesn't that leave documents with no text as essentially invisible?"* If No, then ask this follow-up question: *"If you do run across a relevant non-textual document, how do you find similar documents if you only classify documents using text?"*
- **Identifying and Correcting Document Unitization Issues.** *"Does your system have any way to detect or correct document unitization issues, e.g., Four Authorizations for Expenditures in the same PDF as a well log? What happens in your system to such a compound document?"*
- **1 TB Free Sample, 1 Week Turnaround.** *"One classification vendor has offered a one terabyte sample processed for free with a one week turnaround time, what can you offer so we can do an apples-to-apples comparison from the same data?"*
- **Number of Document Types.** *"How many document types can your system classify? What's the most you've ever done?"*
- **Document Attribute Extraction.** *"What does your system do to help me extract document attributes from documents, e.g., pull the well number off well logs, or pull the name of contracting parties off contracts?"*
- **New Document Type Alert.** *"For use in ongoing information governance initiatives, will your system alert the user if new types of documents start appearing or is it more batch oriented for litigation-type applications?"*
- **Limitations on Offerings or Experience.** *"Does your company offer any document classification technology that isn't text dependent?"* OR for speakers, *"Have you ever worked with a document classification technology that wasn't text dependent?"*

What you may find is that TAR/PC speakers and vendors have made a virtue out of necessity. In other words, for many years the most effective way to deal with large volumes of documents was text, so they used it and got pretty good at it despite the limitations of a text-dependent approach. However, using something because it was at one time expedient is no reason to continue when there is a more comprehensive and consistent approach with visual classification.

See, [TAR Defensibility Soft Spots: Text Dependence and Document Unitization](#) or [Information Governance Lessons from 4 AFEs and a Daily Drilling Report](#)