

Court-Approved Predictive Coding Can Save Discovery Dollars in Data-Intensive Litigation

February 29, 2012

U.S. Magistrate Judge Andrew Peck of the Southern District of New York found himself in the enviable position of granting his own wish. As an active commentator on the ever-changing landscape of e-discovery, Judge Peck had written an article lamenting litigants' reluctance to use predictive coding technology until some court somewhere blessed the technology. A few months later, he issued the first order approving an e-discovery protocol centered around predictive coding technology in [*Monique Da Silva Moore v. Publicis Groupe & MSL Group*](#), Case No. 11-cv-01279 (S.D.N.Y. Feb. 24, 2012).

What is Predictive Coding?

Predictive Coding is a technique in which attorney reviewers interact with computer software to "train" the computer to cull responsive documents from a large body of electronically stored information (ESI). The process begins with small "seed set" of documents taken from the broader universe of a party's ESI. Often, the seed set is identified using keyword searches. The attorney reviewers then code all the documents in the seed set as either responsive or non-responsive. Frequently, the reviewers will also flag "hot" documents in the seed set. Based on a variety of characteristics of the documents coded by the reviewers, the computer then applies algorithms to all of the ESI and returns a much smaller set of documents that are likely to be responsive.

To fine tune the process, the reviewers can use a series of "iterations" in which they review and code additional samples of documents returned by the computer to further "train" the computer on the characteristics of responsive documents. Ultimately, the predictive coding process can reduce a universe of millions of documents to just a few thousand that attorneys need to manually review

before production. When done correctly, this can save substantial discovery costs.

What Did Judge Peck's Opinion in *Moore* Decide?

All the court actually did in the *Moore* opinion was approve a discovery protocol that the parties had largely agreed to after several discovery hearings. During those hearings, Judge Peck had helped the parties overcome the plaintiffs' initial reluctance to cooperate with the defendants' plan to use predictive coding. Along the way, though, the court offered a helpfully detailed discussion of the use of predictive coding, and some guidance that future litigants can provide to less tech-savvy courts.

In *Moore*, a group of female public relations employees sued one of the world's four largest advertising firms alleging gender discrimination. The defendants had an ESI universe of more than 3 million documents. Due to proportionality concerns, the defendants wanted to limit their review cost to about \$550,000. Although the court would not enforce that limit prior to actually engaging in the initial review, it did approve most of the details of the predictive coding process that the defendants proposed. Those details are less significant than the guidance the *Moore* court provided for litigants considering using predictive coding:

- **Cooperation remains the order of the day.** Responding to a press release from the plaintiffs' e-discovery vendor, a number of [commentators](#) have suggested that the *Moore* court actually ordered the parties to use predictive coding. It didn't. The parties agreed to use predictive coding. True, the plaintiffs were reluctant to agree to the defendants' proposed process, were strongly coaxed by the court to agree, and preserved their objections when they did. But the court expressed no opinion on whether it would have ordered predictive coding over the plaintiffs' objections, and the court repeatedly urged parties to work with each other in good faith conference to agree on a protocol.
- **Transparency is your friend.** The *Moore* court also repeatedly commended the defendants for the transparency of their proposed process. The defendants proposed to share with plaintiffs the entire seed set of documents (withholding only privileged documents), along with the initial coding of those documents. They would continue to share that information during each of seven proposed iterations to train the system. That transparency appeared to give Judge Peck the comfort that the defendants' goal was genuinely controlling costs, rather than trying to avoid producing relevant documents.
- **"Bring your geek to work day."** The court also commended both parties for bringing their respective e-discovery vendors to speak to the court during discovery hearings, jokingly calling the practice "bring your geek to work day." The court ruled, however, that such vendors need not be sworn to offer their comments, and that the Federal Rules of Evidence and the *Daubert* standard for expert testimony do not generally apply in the context of discovery hearings like this one. In essence, the court treated the vendors more like counsel or officers of the court, who could make unsworn representations of fact to the court.

- **Predictive coding may be more effective than the “gold standard” of complete review.** Parties opposing the use of “computer-assisted” methods in e-discovery have long asserted that the discovery rules require the “gold standard” of manual review of the entire universe of documents. But the *Moore* court very helpfully summarized a growing body of empirical research demonstrating just the opposite. As it turns out, for large bodies of documents, attorney review seems to be more error prone and underinclusive, not less. Only about five percent of human errors stemmed from genuine disputes about relevance in borderline cases, with fatigue and inattention causing the rest. The court also noted that predictive coding can allow more experienced senior attorneys to play more influential roles in coding, presumably improving the results.

Conclusion

In data-intensive litigation, predictive coding can help substantially reduce discovery costs. If you want it to work for you, take a cooperative and transparent approach, involve your experts, and use the *Moore* opinion for support.

©2024 Carlton Fields, P.A. Carlton Fields practices law in California through Carlton Fields, LLP. Carlton Fields publications should not be construed as legal advice on any specific facts or circumstances. The contents are intended for general information and educational purposes only, and should not be relied on as if it were advice about a particular fact situation. The distribution of this publication is not intended to create, and receipt of it does not constitute, an attorney-client relationship with Carlton Fields. This publication may not be quoted or referred to in any other publication or proceeding without the prior written consent of the firm, to be given or withheld at our discretion. To request reprint permission for any of our publications, please use our Contact Us form via the link below. The views set forth herein are the personal views of the author and do not necessarily reflect those of the firm. This site may contain hypertext links to information created and maintained by other entities. Carlton Fields does not control or guarantee the accuracy or completeness of this outside information, nor is the inclusion of a link to be intended as an endorsement of those outside sites.