

## COMPLEX LITIGATION & *E-Discovery*

### Computer-Assisted Review of E-Discovery

The first order and opinion approving its use; has the future arrived?

---

BY BETH S. ROSE AND  
CHARLES J. FALLETTA

---

**W**ith the costs of e-discovery placing increasing financial burdens on litigants, attorneys have attempted to control e-discovery costs by significantly reducing the quantity of electronically stored information (ESI) collected and manually reviewed before production. One of the ways of reducing ESI is by agreeing to an e-discovery protocol setting forth limitations and parameters for ESI productions,

including limiting the number of custodians, removing duplicates, incorporating claw-back procedures for privileged documents inadvertently produced and using “keywords.” Among other things, an e-discovery protocol can help narrow the scope and reduce the amount of data needed to be collected and reviewed. However, even after the amount of ESI has been narrowed, it still must be manually reviewed and coded by attorneys or paralegals before being produced, often taking hundreds, if not thousands, of hours and at great expense.

What if senior attorneys could review a small subset of the ESI collected and use that minimal review to train a computer to identify and code responsive documents for production? Would a court permit such a protocol? Could computer-assisted coding, also referred to as predictive coding, be as accurate as

attorney or paralegal review? Would the computer-assisted coding save money? At least one judge believes the answer to these questions is “yes” and has issued a decision advocating for parties to use computer-assisted coding in appropriate cases.

That judge is the Honorable Andrew Peck, United States Magistrate Judge, of the Southern District of New York. In the fall of 2011, Judge Peck authored an article, “Search, Forward: Will manual document review and keyword searches be replaced by computer-assisted coding?” *L. Tech. News*, Oct. 2011, lauding the benefits of computer-assisted coding in cases “where it will help ‘secure the just, speedy, and inexpensive’ (Fed. R. Civ. P. 1) determination of cases in our e-discovery world.” Computer-assisted review involves the use of “sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer.” In essence, a computer program is able to determine the relevance of large numbers of documents based on the results of an attorney’s review and coding of a small “seed” set of documents. By using the computer software to predict relevance, attorney hours spent reviewing large quantities of nonrelevant documents are avoided and corresponding savings are passed on to clients.

In his article, Judge Peck further

---

*Rose is chairwoman of the Sills Cummis & Gross product liability practice group and co-chair of the firm’s litigation practice group. Falletta is a member of the firm’s litigation and product liability litigation practice groups. The views and opinions expressed in this article are those of the authors and do not necessarily reflect those of Sills Cummis & Gross PC.*

urged counsel “to rely on [his] article as a sign of judicial approval” of computer-assisted coding “until there is a judicial opinion approving (or even critiquing) [its] use ....” Fortunately for those waiting for such a judicial opinion, the wait was not very long. In February 2012, Judge Peck had the opportunity to author the first opinion and order approving the use of computer-assisted coding, which opinion and order was adopted by the District Court on April 26, 2012, over the plaintiffs’ objections.

In *Da Silva Moore v. Publicis Groupe & MSL Group*, No. 11-1279, 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. Feb. 24, 2012), five plaintiffs asserted individual and class-action gender discrimination claims under federal and New York state law. After several discovery conferences and rulings issued by the district court judge, the matter was referred to Judge Peck for pretrial supervision. At the time of the referral, the parties had been working on an e-discovery protocol and agreed on certain custodians from whom to collect data, but had several unresolved disputes including whether to use computer-assisted coding to reduce the approximately three million electronic documents identified by the defendants.

While the parties eventually agreed that computer-assisted coding should be used, the plaintiffs disagreed with the methodology proposed by the defendants. The defendants proposed that a senior attorney would first review and code 2,399 documents as a “seed” set that would be used to train the predictive coding software. Senior attorneys for the defendants would develop this seed set by first using “judgmental sampling,” that is, by specifically targeting certain custodians and the issues of the case likely to retrieve relevant documents. Next, these senior attorneys would conduct targeted keyword searches, including keywords proposed by the plaintiffs.

In addition, to stabilize the training of the software of the defendants’ e-discovery vendor, the defendants would use seven iterative rounds, during which the software would rank the documents on a scale of zero to 100 (with 100 being the most relevant). Senior attorneys would review at least 500 documents from each of the rounds to see if the software

was returning new relevant documents. After seven rounds, the defendants would review a random sample, not selected by the software, to confirm that the computer was not discarding relevant documents. The defendants further agreed to provide the plaintiffs with all nonprivileged documents reviewed throughout this entire process, irrespective of whether they were coded relevant or irrelevant, so that the plaintiffs could review the manner in which the defendants coded the documents.

Although the plaintiffs’ e-discovery vendor generally agreed that “computer-assisted review works, and works better than most alternatives,” the plaintiffs reserved their right to object and, in fact, filed several objections. The plaintiffs first argued that the court’s acceptance of the defendants’ computer-assisted review protocol provided “unlawful cover” for defendants’ counsel who had a duty under Federal Rule of Civil Procedure (FRCP) 26(g) to certify that their document production was “complete” and “correct.” Judge Peck, however, determined that no attorney could honestly certify that a production involving over three million e-mails was “complete,” and further determined that FRCP 26(g) only applied to parties’ initial disclosures under FRCP 26(a)(1), and not to defendants’ discovery responses.

The plaintiffs next argued that the court’s acceptance of the computer-assisted review protocol was contrary to Federal Rule of Evidence (FRE) 702 and the court’s gatekeeper role in precluding unreliable expert testimony from being submitted to the jury at trial. Judge Peck rejected the plaintiffs’ argument by stating that FRE 702 did not apply to how documents are produced in discovery because the defendants were not seeking to introduce the ESI protocol results into evidence, and that the admissibility of each e-mail produced would be considered at trial on an e-mail-to-e-mail basis.

The plaintiffs further objected on the grounds that there was no way of determining whether the defendants’ protocol was reliable. Judge Peck, however, determined that the plaintiffs’ objection was premature because the concerns they raised had not yet occurred and were better decided “‘down the road’ when real

information is available to the parties and the Court.” In overruling the plaintiffs’ objections, Judge Peck determined that computer-assisted coding was appropriate in that case because: (1) the parties agreed to use it; (2) the amount of ESI exceeded three million documents; (3) computer-assisted review was superior to manual review or keyword searches; (4) FRCP 26(b)(2)(C) required cost effectiveness and proportionality; and (5) the proposed protocol was transparent.

Judge Peck also made several observations about the benefits of computer-assisted coding. First, he noted that “computer-assisted review is not a magic, Staples-Easy-Button, solution appropriate for all cases,” but that the “[t]echnology exists and should be used where appropriate.” Second, Judge Peck emphasized that the objective of e-discovery review is to “identify as many relevant documents as possible, while reviewing as few nonrelevant documents as possible,” and that “statistics clearly show that computerized searches are at least as accurate, if not more so, than manual review.” Third, Judge Peck was critical of keyword searches alone, stating that, while they have their role in locating relevant ESI, there are inherent problems with their accuracy and efficiency: “In too many cases, however, the way lawyers choose keywords is the equivalent of the child’s game of ‘Go Fish.’ The requesting party guesses which keywords might produce evidence to support its case without having much, if any, knowledge of the responding party’s ‘cards’ (i.e., the terminology used by the responding party’s custodians).” In addition, Judge Peck indicated that keyword searches often return large quantities of “false positive” results requiring the need for expensive and labor-intensive manual reviews.

Lastly, Judge Peck noted that while computer-assisted coding is not perfect, it is better than other available alternatives and that the FRCP do not require perfection. Instead, the FRCP are aimed at securing the “just, speedy, and inexpensive determination” of disputes, while balancing the proportionality requirements that discovery not be unreasonably cumulative, duplicative, or require production when it is available from more convenient and less burdensome or expensive sourc-

es. (Citing FRCP 1 and 26(b)(2)(C).)

In cases where litigants are faced with large quantities of ESI, consideration should be given to incorporating into an e-discovery protocol the use of computer-assisted predictive coding technology,

along with other available e-discovery tools, to help reduce costs and increase the efficiency and accuracy of ESI production. Because the *Da Silva Moore* case appears to be the first decision approving the use of computer-assisted review, it is

an important decision to be utilized by counsel in future ESI intensive litigations. In that regard, counsel will be happy to know that the court attached to its decision a copy of the parties' e-discovery protocol that could be used as a model. ■