

# The Legal Intelligencer

Reprinted with permission from the 06-27-2012 issue of The Legal Intelligencer. (c) 2012 ALM Media Properties.

Further duplication without permission is prohibited.

---

## Predicting the Future of Predictive Coding

A decade ago, document review meant a small militia of lawyers sitting in a windowless warehouse surrounded by bankers' boxes full of paper documents.

Hayes Hunt and Jillian R. Thornton

2012-06-27 12:00:00 AM

A decade ago, document review meant a small militia of lawyers sitting in a windowless warehouse surrounded by bankers' boxes full of paper documents. Now, thanks to extreme information inflation, the bulk of document review takes place electronically. In order to keep up with the enormous volume of electronically stored information, lawyers have employed a method featuring a combination of keyword searches and manual review. Most importantly, e-discovery can be responsible for 70 to 90 percent of the client's cost of litigation. However, recently, the universe of ESI has expanded in exponential fashion. Exabytes have devoured the smaller gigabytes in the ESI pond. What's next? Predictive coding.

Predictive coding is being used to run algorithms that allow for computer characterization of a massive set of electronic data for a fraction of the cost of more traditional methods (i.e., a cadre of lawyers). Case law is now catching up to the technology and various judges are giving the green light for lawyers to employ predictive coding in e-discovery without running afoul of the rules. The proper use of predictive coding, especially in large-data-volume cases, provides huge benefits for lawyers and clients: Predictive coding of ESI takes much less time, saves a lot of money and is often as accurate or more accurate than manual review. Of course, predictive coding also can be problematic if, for example, privileged documents are disclosed.

A recent study by Rand Corp., which includes 57 case studies from eight large corporations, shows that the cost of e-discovery can be grouped into three main categories: collection, processing and review. Amazingly, the review phase accounted for 73 percent of the costs incurred during e-discovery. Predictive coding works to drastically reduce the number of documents that are manually reviewed by lawyers. Here's how it works: The first step in the process is that lawyers review a small sample of documents and code those documents for relevance or privilege or subject matter. The software then studies the sample set and applies the coding principles that it has learned to a larger set of documents. Then, the lawyers review the computer-coded documents to further teach the program how to code. This program continues until the software identifies only relevant documents. After coding is finished, the software can be used to select a small, random population of documents for lawyers to perform quality-control checks. If errors are found, the lawyers code more sample documents until accuracy of the coding reaches an acceptable level. Then the review is complete. The software can reduce the documents that need to be manually reviewed from a set of 2 million, for example, to only 3,000 to 5,000 documents. Assume it takes a lawyer 60 seconds to review a one-page document and you can easily do the cost-effective math of predictive coding.

## What Judges are Doing with Predictive Coding

The difficulty with implementing processes such as predictive coding is that the technology is so new that these methods are fairly untested in court. However, the first wave of cases discussing the propriety of predictive coding has illustrated that it is indeed going to be accepted as an appropriate discovery tool. Earlier this year, in a landmark decision, U.S. Magistrate Judge Andrew J. Peck for the Southern District of New York authorized the use of predictive coding in *Da Silva Moore v. Publicis Groupe*, No. 11-CV-1279, 2012 U.S. Dist. LEXIS 23350 (S.D.N.Y. Feb. 24, 2012). Peck summarized his position, stating: "What the bar should take away from this opinion is that computer-assisted review is an available tool and should be seriously considered for use in large-data-volume cases where it may save the producing party (or both parties) significant amounts of legal fees in document review." The district court ultimately adopted Peck's evidentiary rulings in *Da Silva Moore v. Publicis Groupe*, No. 11-CV-1279, 2012 U.S. Dist. LEXIS 58742 (S.D.N.Y. Apr. 26, 2012). In his opinion, Peck was careful to point out that the plaintiffs consented to the defendant's use of predictive coding and the discovery dispute merely concerned the implementation of its use. Thus, lawyers were left to wonder what would happen if the parties did not agree to the use of predictive coding.

The same week that Peck's ruling was affirmed, a state court judge in Virginia approved the use of predictive coding in a case over the objections of the opposing party. In *Global Aerospace v. Lindow Aviation*, No. CL 61040 (Vir. Cir. Ct. Apr. 23, 2012), 20th Judicial Circuit Judge James Chamblin ordered that defendants could use predictive coding, despite the plaintiff's objections that the technology was not as effective as manual review. Chamblin disagreed and ordered the predictive coding for the production of the defendant's ESI, provided that the receiving party would still have an opportunity to question the completeness of the contents of the production or the ongoing use of predictive coding. This opinion, although limited in its direct impact in other litigation, along with Peck's decision in *Da Silva Moore*, indicates willingness by the judiciary to incorporate predictive coding into e-discovery.

Predictive coding and other automated methods of e-discovery obviously have limitations. Peck, in *Da Silva Moore*, emphasized that his approval of predictive coding was not universal: he did not order the use of predictive coding, he stated that computer-assisted review is not required in all cases, and that he did not endorse any particular vendor or predictive coding tool. In addition, it is the responsibility of the lawyers to understand the predictive coding program and how it works so that they can demonstrate the method's reasonableness if it is called into question. Just as with the traditional e-discovery tool of keyword searching, lawyers must engage in careful planning and sufficient quality control to ensure the accuracy of the program. Finally, lawyers should cooperate with opposing counsel and be transparent in their use and the scope of predictive coding in order to avoid unnecessary discovery disputes.

## Clawing Back Privileged Documents

With regard to privilege review, lawyers utilizing predictive coding of ESI need to be especially vigilant not to inadvertently produce privileged documents. Although predictive coding can be used to assess privilege as well as relevance, lawyers need to evaluate the benefit compared to the risk of disclosure. Under the federal rules governing clawback, a "disclosure of a communication or information covered by the attorney-client privilege or work product protection ... does not operate as a waiver in a federal or state proceeding if (1) the disclosure is inadvertent; (2) the holder of the privilege took reasonable steps to prevent disclosure," as well as reasonable steps to correct the error. Thus, whatever method of ESI review lawyers use, it must rely on reasonable steps to prevent disclosure. This is a subjective standard, and lawyers using predictive coding would be wise to carefully document the process of how they code for privilege. In at least one federal case, the court has held that a party waived its right to attorney-client privilege by mistakenly producing privilege documents after employing a faulty keyword filter. See *Victor Stanley v. Creative Pipe*, 250 F.R.D. 251 (D. Md. 2008). Part of the court's decision was based on the defendant's "regrettably vague" explanation of how the keywords were developed, how the search was conducted and what quality controls were employed. Thus, the need for precision in designing the search program and extensive quality control is obvious.

Given the ever-expanding universe of ESI, most lawyers would be wise to consider a review and especially predictive coding. After all, the research has shown that predict

precise, makes fewer errors and identifies more relevant documents than human reviewers. This should not come as a surprise when one considers the differences in opinion among lawyers about what information is "relevant." When you add millions of pages of documents, fatigue plays a role for human reviewers. Based on these factors in addition to the dramatic saving of time and money, it is clear that predictive coding and similar methods are going to revolutionize how we conduct e-discovery. •

**Hayes Hunt**, a member of **Cozen O'Connor** in the firm's commercial litigation and criminal defense and government investigations practice groups, concentrates his practice in the representation of individuals, corporations and executives in a wide variety of federal and state criminal law and regulatory enforcement matters, as well as complex civil litigation. He can be reached at [hhunt@cozen.com](mailto:hhunt@cozen.com).

**Jillian R. Thornton** is an associate in the firm's commercial litigation group. She focuses her practice on securities and shareholder disputes and complex commercial litigation.

---

Copyright 2012. ALM Media Properties, LLC. All rights reserved.