

# Electronic Discovery Problems in Employment Litigation

---



**Lynne Bernabei** has been litigating employment discrimination, civil rights, and whistleblower protection cases for over thirty years. She is a founding partner of Bernabei & Wachtel, PLLC where she handles claims of civil rights violations, discrimination, and other employment and contract disputes. Ms. Bernabei has developed extensive trial experience in federal and local courts in the District of Columbia, Maryland

and Virginia and before administrative agencies. Ms. Bernabei has argued appeals before the United States Courts of Appeals for the Third Circuit, the Fourth Circuit, the Sixth Circuit, the D.C. Circuit, and the Federal Circuit. Ms. Bernabei serves as a member of the editorial board of the *Labor Law Journal* and formerly served on the board of the *Journal of Employment Discrimination*. She is also a faculty member of the American Law Institute-American Bar Association and teaches courses on employment law and civil rights matters for the District of Columbia Bar and as a guest lecturer at law schools. Ms. Bernabei has been selected as a member of the College of Labor and Employment Lawyers and is also a member of the National Employment Lawyers Association and the American Bar Association. Ms. Bernabei has been approved by the American Arbitration Association (AAA) for its National Roster of Neutrals.



**Alan Kabat** received a Bachelor of Arts degree from the University of Washington in 1983. He received a Ph.D. in Biology from Harvard University in 1990. Mr. Kabat received a J.D. degree from Georgetown University Law Center in 1998. He was a member of the American Criminal Law Review, and served as Editor-In-Chief for its Thirteenth Survey of White Collar Crime. Mr. Kabat is a member of the American Bar Association and the Metropolitan Washington

Employment Lawyers' Association (MWELA). He is currently co-chair of the Amicus Committee for MWELA. He also serves as discovery coordinator for a complex multidistrict litigation matter. Mr. Kabat has practiced law with Lynne Bernabei since 1998. Mr. Kabat has been admitted to practice law in Maryland since 1998, in the District of Columbia since 1999, and in Virginia since 2008. He is also admitted to practice before the U.S. Supreme Court; the U.S. Court of Appeals for the District of Columbia, Second, Fourth, Sixth, Ninth, and Tenth Circuits; and the U.S. District Courts for District of Columbia, Maryland, and the Eastern District of Virginia.

---

## Lynne Bernabei and Alan R. Kabat

---

**There are more things to discover in today's employment cases—and more ways to discover them.**

---

**THIS ARTICLE** discusses several recent issues in electronic discovery in employment litigation: (i) the use of keyword searching versus predictive coding for reviewing and producing large sets of documents; (ii) privilege and waiver issues arising from electronic discovery, including the impact of Fed. R. Evid.502; and (iii) ethical issues arising from accessing another party's emails or social media sites.

### **PREDICTIVE CODING AND KEYWORD SEARCHES FOR VOLUMINOUS DOCUMENTS**

• In the “dark ages”—only a decade or so ago—the customary practice in reviewing documents in order to identify the relevant, responsive, and non-privileged documents for production was to have an attorney or paralegal review each page of each document to determine whether: (i) it should be produced; or (ii) withheld in part or entirely on the grounds of privilege; or (iii) not produced as not responsive. As the

volume of documents grew exponentially, thanks to the proliferation of email, attorneys and their clients realized that this manual, page-by-page search, could be expensive and time-consuming, particularly in class actions that might involve hundreds or thousands of employees working in multiple offices.

Hence, attorneys started doing keyword searches and searches limited to certain record custodians. The parties would agree to limit the document searches to those in the electronic files and emails of a defined subset of employees (“custodians”), and that those documents would be electronically searched by keywords. Thus, only the documents that came up through a keyword search would have to be reviewed for privilege, with the presumption being that any document with one or more keywords would be relevant and responsive, while documents without any keywords would not be relevant or responsive and need not be considered any further. Keyword searching, of course, assumes that the parties can reach agreement on the keywords and the record custodians. Keyword searching can also be combined with “Boolean” searching, which is similar to searches done in Westlaw or Lexis, where one uses logical connections (*e.g.*, “and” “or” “w/10”) to search for court decisions containing certain words or phrases, including those are close to each other, as opposed to court decisions in which those words are widely separated.

For example, in employment litigation, the records custodians could be defined to include the following: (i) plaintiff(s); (ii) the supervisors and other employees involved in the alleged adverse employment actions; (iii) comparators; and (iv) co-workers who are known or believed to have witnessed the employment actions.

Similarly, keywords in employment litigation would include not only the names of the plaintiff(s)

and any individual defendant(s), in order to capture emails in which *other* people discussed the parties, but also words relating to important terms in the complaint and the answer. For example, in a case involving discrimination relating to bonus payments, keywords would include the name of the company’s bonus or incentive program, as well as the terms used for the targets in that program. In a case involving discrimination based on a failure to provide comparable training opportunities to all employees, keywords would include words and phrases relating to the company’s training programs and the selection process for those programs. The parties can also conduct staged searches, starting with a subset of keywords, and using the results to identify further keywords to be used, as a way of more precisely focusing the search.

However, there has been an increasing recognition that keyword searching is not a panacea. It is both overinclusive and underinclusive. For example, searching for emails with an employee’s name may yield emails about other employees with the same first and/or last name—overinclusive. Conversely, searching for emails in which stereotyped remarks are made will not yield emails that use phrases that the attorneys did not think of including in the keywords—underinclusive. These problems led U.S. Magistrate Judge Andrew Peck to conclude that keyword searching is not the last word in electronic discovery:

“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). Indeed, the responding party’s counsel often does not know what is in its own client’s ‘cards.’”

*Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182, 191 (S.D.N.Y. 2012), *adopted by*, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012).<sup>1</sup>

Similarly, U.S. District Judge Shira Scheindlin recently concluded that: “There is increasingly strong evidence that ‘keyword searching is not nearly as effective at identifying relevant information as many lawyers would like to believe.’” *National Day Laborer Organizing Network v. U.S. Immigration & Customs Enforcement Agency*, 877 F. Supp. 2d 87, 109 (S.D.N.Y. 2012) (quoting Maura R. Grossman & Terry Sweeney, “What Lawyers Need to Know About Search Tools: The Alternatives to Keyword Searching Include Linguistic and Mathematical Models for Concept Searching,” *National Law Journal* (Aug. 23, 2010)).

Thus, the latest wave in e-discovery is “predictive coding,” also known as “computer-assisted coding.” Magistrate Judge Peck, who authored the first reported decision on predictive coding, noted that most lawyers already use predictive coding without being aware of it—the “spam filters” in email programs use predictive coding to segregate emails that are presumed to be spam, *i.e.*, pornographic emails, advertisements and solicitations from sources known or likely to be scams, and other emails of questionable veracity. *Da Silva Moore*, 287 F.R.D. at 184 n.3 (spam filters use predictive coding). Unfortunately, attorneys are well aware that spam filters are both overinclusive and underinclusive—legitimate client email sometimes gets trapped in the spam filters, while true spam shows up in inboxes. Nonetheless, predictive coding is here and employment lawyers need to prepare for it.

<sup>1</sup> The plaintiffs in *Da Silva Moore*, dissatisfied with Magistrate Judge Peck’s rulings, sought to disqualify him based on the fact that he wrote an article about predictive coding, so that he was allegedly biased in favor of predictive coding. The court denied their motion, *see* 868 F. Supp. 2d 137 (S.D.N.Y. 2012), and the Second Circuit denied their mandamus petition. *In re Da Silva Moore*, No. 12-5020 (2d Cir. Apr. 10, 2013) (*per curiam*).

Magistrate Judge Peck provided a general definition of this process: “By computer-assisted coding, I mean tools (different vendors use different names) that use sophisticated algorithms to enable the computer to determine relevance, based on interaction with (i.e., training by) a human reviewer.” *Id.* at 183-84.

Unlike manual review, where the review is done by the most junior staff, computer-assisted coding involves a senior partner (or [small] team) who review and code a “seed set” of documents. The computer identifies properties of those documents that it uses to code other documents. As the senior reviewer continues to code more sample documents, the computer predicts the reviewer’s coding. (Or, the computer codes some documents and asks the senior reviewer for feedback.)

When the system’s predictions and the reviewer’s coding sufficiently coincide, the system has learned enough to make confident predictions for the remaining documents. Typically, the senior lawyer (or team) needs to review only a few thousand documents to train the computer.

Some systems produce a simple yes/no as to relevance, while others give a relevance score (say, on a 0 to 100 basis) that counsel can use to prioritize review. For example, a score above 50 may produce 97 percent of the relevant documents, but constitutes only 20 percent of the entire document set.

Counsel may decide, after sampling and quality control tests, that documents with a score of below 15 are so highly likely to be irrelevant that no further human review is necessary. Counsel can also decide the cost-benefit of manual review of the documents with scores of 15–50. *Da Silva Moore*, 298 F.R.D. at 183 (quoting A. Peck, *Search, Forward*, L. Tech. News, Oct. 2011, at 25, 29); *see also* Maura R. Grossman & Gordon V. Cormack, “Glossary of Technology-Assisted Review,” 7 *Fed. Cts. L. Rev.* 1 (2013) (technical definitions of the terms used in this field) (online at: <http://www.fclr.org/fclr/articles/html/2010/grossman.pdf>).

How exactly does predictive work, putting aside proprietary software methods? The following description is taken from *Da Silva Moore*, an employment case brought against a large advertising agency, in which the named plaintiffs sought to bring both a class action and a collective action under federal and New York anti-discrimination statutes. The employer had some three million electronic documents that had to be reviewed for responsiveness, relevancy, and privilege. Defendant, over plaintiffs' objection, insisted on pursuing predictive coding, thus resulting in the first reported opinion in this field.

First, the parties must agree upon record custodians, temporal scope of discovery, and upon the broad categories, keywords, or "issues" to be used to identify relevant documents.

Then, the producing party selects a random sample (*e.g.*, several thousand documents) of the entire email and other electronic files associated with those record custodians.

An experienced attorney for the producing party, familiar with the claims and defenses in the lawsuit, then reviews that subset, and "codes" them as to whether they are relevant, and as to which of the issue(s) are associated with each document. The result is that each document in the subset is scored with a code (on a 1-100 scale) as to relevancy. The attorney also sets aside documents that are privileged.

The software program keeps track of how each document is coded by the attorney. Based upon that data, the program uses proprietary technology to develop algorithms for analyzing the pool of documents to identify and code the most relevant documents.

The producing party then turns over the non-privileged subset of the reviewed documents, together with the code for each such document, and allows the receiving party to do its own predictive coding analysis of that subset of documents. If the parties reach significantly different results as to what

is relevant, they can discuss the coding process, and do successive iterations, until both sides are satisfied with the coding of each document in that subset. The parties also have to agree on the cut-off level for the codes, *i.e.*, the number on the scale from 0 to 100 for which documents below that cut-off will not be produced.

Then, the software program takes over—based on its collective analysis of the attorneys' coding of the subset, the software then runs through the remaining thousands or millions of documents to code them. The producing party then only has to do a privilege review of the documents above the cut-off level, in order to redact or withhold entirely any privileged documents. Critically, documents that are below the cut-off level will *not* be reviewed by anyone, since they are presumptively not relevant or responsive.

It is obvious that there can be disputes at each stage of this process—the identification of the record custodians; the temporal scope of discovery; the identification of issues and keywords to be used in coding documents for relevancy; the size of the random sample to be analyzed by the attorney; the numerical codes to be assigned to any given document; the number of iterative reviews, if any; and the cut-off score to be used. Indeed, in *Da Silva Moore*, the parties disputed each of these issues. *Da Silva Moore*, 287 F.R.D. at 184-87. Commentators noted that "in some cases, the inability of parties to agree on predictive coding has brought discovery to a standstill and forced courts to weed through infinite discovery motions." *See* Pooja Nair & Leslie Nash Tookey, "Five Cases Made for a Momentous Year in Predictive Coding," *National Law Journal*, Jan. 28, 2013, at 10, 11.

Magistrate Judge Peck noted that his decision "does not mean [that] computer-assisted review must be used in all cases, or that the exact ESI protocol approved here will be appropriate in all future cases that use computer-assisted review." *Da Silva Moore*, 287 F.R.D. at 193. However, the take-away