



HOW TO STOP WORRYING AND LOVE

# PREDIC

In big-data cases, predictive coding can help both plaintiffs and defendants conduct electronic discovery quickly and cost-effectively. With early collaboration between parties and post-review testing, you can be confident you're getting the **relevant documents**.

# PREDICTIVE CODING

By || **ANNIKA K. MARTIN**

Predictive coding is quite the hot topic these days. But what is it? And why should you care? Put simply, it's document review technology. Predictive coding is a type of technology-assisted review<sup>1</sup> that enables a computer, relying on input from a human reviewer, to predict how documents should be classified. Websites and services we use every day, such as Pandora, Amazon, and Netflix, already use similar technology to predict our choices. Just as Amazon uses input from you—your searches, your views, your purchases, your product ratings—to predict what you might like next, predictive coding software uses input from a lawyer's coding of a small set of documents to predict how that lawyer would code the rest of the documents.

The process of using predictive coding software to harness the lawyer's judgments on a smaller set of documents and extrapolate them to the rest of the universe of documents is called predictive review. Once predictive review is complete, the recall (comprehensiveness) and precision (accuracy) of the predictive technology's review can be measured.

Plaintiff counsel should care about this technology because it provides unprecedented insight into the quality of discovery-produced documents. It is faster, cheaper, and more accurate than traditional review, helping us find those "hot docs" more easily.

Predictive coding is a means to a variety of ends. It helps identify relevant documents for production, categorize documents relevant to particular issues, and prioritize review. Predictive coding can—and should—be used in appropriate large-data-volume cases to reduce the cost of document review while increasing speed and accuracy.

This technology can provide insight and input into defendants' document production processes, increase and simultaneously measure the accuracy of defendants' pre-production document review, and result in faster document productions composed almost entirely of relevant documents—which in turn reduces time and money spent finding key documents.

Although some attorneys resist predictive coding, there are few valid arguments against its use in appropriate cases, and courts increasingly favor it.<sup>2</sup> In fact, embracing predictive coding may be more than just a good idea—recent updates to the Model Rules of Professional Conduct obligate plaintiff attorneys to educate themselves about new technologies and to encourage their use when their clients can benefit.<sup>3</sup>

### Insight and Input into the Search Process

**T**raditionally, during discovery, the requesting party had little to no insight into the responding party’s search for, or pre-production review of, documents responsive to discovery requests. Hard-copy files were organized by custodian or by topic and required manual, individual review—there simply weren’t other ways to search and review.

As paper has given way to electronically stored information (ESI), search and review methods have multiplied. The sheer volume of ESI also raises the question of what to search for—words, names, concepts—in all that data. But determining what to search for increases the risk of missing something. If search parameters are not carefully selected, one risks overlooking entire swaths of relevant documents.

In essence, technology is a universal enhancer—it amplifies efficiency and accuracy, but it also amplifies error. Recognizing this double-edged sword—as search technologies have gained ground—courts have increasingly begun to encourage or even require parties to collaborate on the discovery process.<sup>4</sup> This kind of collaboration—working together to choose search terms or custodians—shifts the identification and resolution of document search issues to the front end of the process, so that costly errors can be avoided.

Using predictive coding furthers this trend by also shifting pre-production document review issues to the front end of the process. By giving the requesting party a say in the software’s training, as well as feedback about the recall and precision of its review, predictive coding offers plaintiffs and courts increased confidence, supported by objective statistical evidence, that the defendant’s production is as complete and accurate as possible.

### Accuracy and Consistency

**R**ecent research reveals the astounding inaccuracy and inconsistency of traditional manual document review. In one study, human reviewers missed between 20 percent and 75 percent of all relevant documents, and 90 percent of those mistakes resulted from inarguable human error.<sup>5</sup> Another study found that when one reviewer determined a document to be relevant, there was only about a 50 percent chance that a second reviewer would agree.<sup>6</sup>

In the past, manual review was the gold standard because it was the only option, and there were no metrics to gauge its accuracy or completeness. But mistakes happen even if they are not caught—the difficulty of measuring a manual review’s accuracy does not mean it was error-free. In contrast, predictive coding technology allows you to objectively evaluate its performance and make adjustments to improve it if necessary.

While predictive coding is a marked improvement over manual review, it is not perfect. Predictive review will sometimes miss a hot document or code junk as relevant—but you can mitigate this risk by carefully considering the appropriate algorithm and protocol for your case and by employing smart quality control during the review and validation methods at the end. Perfection is not the goal. As one district court observed, “the idea is not to make [the review process] perfect, it’s not going to be perfect. The idea is to make it significantly better than the alternatives without nearly as much cost.”<sup>7</sup>

Predictive coding cannot—and should not—entirely replace people in the search and review process. In fact, the software-training process actually requires *more* human interaction than other technologies, such as keyword search. While predictive coding effectively transfers the drudgery of review

to the machine, the judgment remains entirely the lawyer’s. Predictive coding is better viewed as a complementary collaboration between people and machines, allowing each to do what they do best.

### Streamlining and Allowing Flexibility

**A**ccording to the RAND Institute for Civil Justice, predictive coding can reduce the cost of document review by as much as 75 percent.<sup>8</sup> These time and cost savings benefit the producing party, but predictive coding can save plaintiffs money as well. If a defendant uses predictive coding for pre-production document review, plaintiffs can get documents sooner, get fewer irrelevant documents, and in some instances even receive documents already issue coded into subject matter categories. This makes it easier and cheaper to find the hot documents necessary to move toward resolution. It essentially sets you up to fish a stocked pond.

Predictive coding also adds flexibility to document review. If search criteria are added or adjusted midway through the discovery process, the lawyer can quickly retrain the software according to the new criteria and rerun the review, at far less cost or time than manual re-review would require.

### Best Practices

**F**irst, determine whether predictive coding is appropriate for your case. This depends on the type of ESI you are reviewing as well as the kind of information you are looking for. Predictive coding is best for cases in which one or both parties will have to search through large volumes of ESI during discovery—either for the pre-production relevance and privilege search or for a topic-based review following the opposing party’s production.<sup>9</sup> Because it’s a



Predictive coding is better viewed as a  
complementary collaboration between people  
and machines, allowing  
**EACH TO DO WHAT THEY DO BEST.**

text-based technology, predictive coding may be unsuitable for identifying relevant handwritten, foreign language, or non-text documents (images, graphics, charts, and spreadsheets) although this is changing as technology advances.

If predictive coding would benefit your case, you should proactively encourage its use early on, ideally as part of the discussions preceding the Rule 26 case management conference. Earlier is better since it can be difficult to change the e-discovery protocol once discovery is under way.<sup>10</sup>

**Know your stuff—or hire someone who does.** You don't have to become a predictive coding expert, but you should learn enough to know when you need one. You can hire an in-house e-discovery technology expert who will stay up to date on new technologies, or you can hire an outside consultant—several firms specialize in e-discovery

consulting for the plaintiff bar.

Expertise is key because you must be able to evaluate the defendants' e-discovery vendors, and you may want to select your own vendor for predictive reviews of large document productions. When selecting an e-discovery vendor or consultant, make sure to do your research, ask around, and comparison-shop to find one with the right skill set and cost profile for your case and budget.

Some defendants and defense counsel will already be familiar with, or even proficient in, predictive coding. Many large corporations already use advanced data management technology in the regular course of business and are keen to use it in litigation as well. The challenge is to level the playing field and avoid being disadvantaged by the defendant's experience.

The best approach may be finding an e-discovery expert who can bring you up

to speed on the technology while also offering advice as you negotiate and implement the e-discovery protocol. Many e-discovery vendors offer stand-alone consultant and expert services for precisely these situations.

Courts also appreciate e-discovery experts' attendance and participation at ESI-related hearings, and judges rely on them to explain complicated e-discovery concepts.<sup>11</sup>

**Get defendants to agree to use it.** Ideally, both parties will reach an agreement early in the litigation about the e-discovery tools and protocol they will use, and will jointly present that discovery plan to the court for approval.<sup>12</sup> If a producing party objects to using a particular method to conduct its pre-production search and document review, a court will rarely force it to do so.<sup>13</sup> Even so, some courts are growing impatient with lawyers refusing to learn about and use available technology.<sup>14</sup>

Plaintiff counsel should work on convincing defense counsel to agree to use predictive coding—another reason to start the e-discovery discussion early.<sup>15</sup> You can raise many arguments in favor of predictive coding. It's faster and cheaper than traditional review, and more likely to follow Federal Rule of Civil Procedure 1, which encourages “speedy” and “inexpensive” determination of every action. If a defendant resists predictive coding—a faster, more cost-effective method of document search and review—it may be unable to certify in good faith that its responses to discovery were conducted in the reasonable manner that Rule 26(g) requires.<sup>16</sup>

Moreover, courts recognize the benefits and risks technology poses in searching and reviewing ESI, and thus most e-discovery protocols incorporate some way to objectively validate the results. If the defendant balks at using predictive coding, ask it, and the court, how it will objectively validate its pre-production



search and review under its preferred method.

Defense counsel may fear the risk of waiving privilege through inadvertent disclosure. However, this concern should be assuaged by recently updated language in Federal Rule of Evidence 502, which provides that inadvertent disclosure of privileged information does not operate as a waiver if, among other things, the disclosing party took “reasonable steps” to prevent the disclosure. Use of predictive coding likely counts.<sup>17</sup>

**Use it as part of a collaborative e-discovery process.** When putting together an e-discovery plan, it is crucial to build transparency and collaboration into the process—in fact, courts require it.<sup>18</sup> Some jurisdictions have even created e-discovery guidelines requiring

parties to collaborate on ESI protocols and other e-discovery matters.<sup>19</sup>

Nevertheless, defendants increasingly push back against these collaboration and transparency requirements, resisting disclosure of key details and denying plaintiff counsel any input into their predictive coding processes. Plaintiff counsel must stand firm and insist on cooperation throughout the process, even including a motion to compel cooperation if needed.<sup>20</sup> Because technology-assisted review inherently risks greatly amplifying mistakes, it is reasonable to require collaboration on performance goals, results validation, and transparency with opposing counsel. Also, statistically validating predictive coding review results is not burdensome in any way. It merely requires transparency—explaining the workflow; disclosing


testing, training, and validation documents within the process; and generating recall and precision reports.

Any e-discovery plan should include many opportunities for collaboration during the process. It is especially important to seek out and use any knowledge your client might have about the defendant’s key custodians and potentially relevant document sources. Then you can approach e-discovery discussions with suggestions for potential ESI sources and keywords.<sup>21</sup> This can be particularly useful in cases where plaintiffs have knowledge about the location of potentially responsive ESI, such as in employment cases.

The software should be trained through a transparent and collaborative process. For example, the ESI protocol could specify that once defense counsel has reviewed and coded a particular training set, you will have the opportunity to review the coding of all non-privileged documents in that set, and you will have the right to challenge and correct defendants’ coding decisions if necessary.<sup>22</sup> You must have input at the training stage, because a predictive coding algorithm’s results are only as good as its training—if the lawyer’s training input is not accurate, the technology will amplify that error, replicating it throughout the entire review. If you have input during the training stage, you have the chance to catch mistakes—or disputes—before they multiply.

The parties and their experts or vendors should also agree on what constitutes satisfactory recall and precision scores that will validate the review’s effectiveness. While setting a numerical target for recall and precision scores may be difficult in some instances, parties can at least agree on a broader validation process initially, and then as review concludes, discuss the recall and precision scores in the context of the review’s size, the prevalence of relevant documents,

## DISABILITY INSURANCE DENIALS



W

MARC WHITEHEAD  
& ASSOCIATES

Attorneys at Law, L.L.P.

Marc Whitehead

American Association for Justice- Leaders Forum  
Board Certified-Personal Injury Trial Law  
Texas Board of Legal Specialization  
Board Certified- Social Security Disability Law  
National Board of Trial Advocacy  
AAJ- Insurance Law Section- Chair Elect  
AV Preeminent Rated Martindale-Hubbell

5300 MEMORIAL, SUITE 725  
HOUSTON, TEXAS 77007  
888-699-8210  
marc@marcwhitehead.com

**O**ur Firm represents sick or injured workers who have been denied their Long Term Disability Insurance benefits. We handle claims in all 50 states. We specialize in the following:

Long Term Disability Insurance Denials

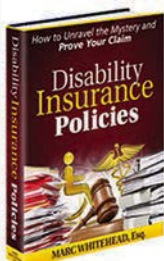
- Administrative Appeals
- Federal ERISA Claims
- Bad Faith Insurance Denials

Fighting Long Term Disability Insurance Denials

- Unum, Cigna, Aetna
- The Hartford, Liberty Mutual
- Prudential, MetLife & More

Referral Fees Paid in Appropriate Cases


- Co-Counsel Arrangements  
Let us do the heavy lifting for you
- Email me to sign up for *The Successful Barrister* Attorney Newsletter!



Visit our website  
to download  
our FREE LTD E-book!  
[www.DisabilityDenials.com](http://www.DisabilityDenials.com)

and other case-specific factors to determine whether review was satisfactory.

A good ESI protocol is also a valuable hedge against defensibility questions the court may raise regarding a proposed e-discovery plan that includes predictive coding. The touchstone here is Rule 26(g)'s "reasonableness" requirement, which, in the context of technology-assisted review methods, courts have interpreted to require transparency and collaboration on the front end, and statistical validation and proof of efficacy on the back end.<sup>23</sup>

E-discovery protocols will vary significantly depending on the case and the type of ESI at issue. No two ESI protocols will be the same—nor should they be. It is critical to consider the unique circumstances of your case when deciding the details of your ESI protocol.<sup>24</sup> Predictive coding can offer you and your clients many benefits. It can help plaintiffs get discovery from defendants more quickly, with productions heavy on wheat and light on chaff, and—thanks to those objective performance metrics—you can be more confident you've gotten everything you asked for. And in your own reviews of documents produced, it can help you find the hot documents faster and at less cost. 



**Annika K. Martin** is a partner at Lieff Cabraser Heimann & Bernstein in New York City. She can be reached at [akmartin@lchb.com](mailto:akmartin@lchb.com).

#### NOTES

1. Predictive coding is sometimes referred to as TAR, but there are other types of TAR besides predictive coding. See Maura R. Grossman and Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review*, 7 Fed. Cts. L. Rev. 1, 26 (2013).
2. See, e.g., *Moore v. Publicis Groupe & MSL Grp.*, 287 F.R.D. 182, 191 (S.D.N.Y. 2012) ("Computer-assisted review appears to be better than the available alternatives, and thus should be used in appropriate cases.").
3. See Model Rules of Prof'l Conduct R. 1.1 cmt. 8 (2012) ("To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology. . .").
4. See, e.g., *William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co.*, 256 F.R.D. 134, 136 (S.D.N.Y. 2009).
5. See, e.g., Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, XVII Richmond J.L. & Tech. 11, 37–43 (2011), available at [jolt.richmond.edu/v17i3/article11.pdf](http://jolt.richmond.edu/v17i3/article11.pdf).
6. Herbert L. Roitblat et al., *Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review*, 61 J. Am. Soc'y Info. Sci. & Tech. 1 (Jan. 2010).
7. *Moore*, 287 F.R.D. at 187.
8. Nicholas M. Pace & Laura Zakaras, *Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery*, RAND Corp. (2012).
9. See, e.g., *Moore*, 287 F.R.D. at 192.
10. See, e.g., *Kleen Prods. LLC v. Packaging Corp. of Am.*, 2012 WL 4498465 (N.D. Ill. Sept. 28, 2012) (refusing plaintiffs' demand that defendants begin to use predictive coding, where discovery was well under way and defendants had already invested over \$1 million to review and produce documents using other advanced search and analysis technologies).
11. See, e.g., *Moore*, 287 F.R.D. at 193.
12. *Id.* at 184; see also The Sedona Conference, *The Sedona Principles: Best Practices Recommendations & Principles for Addressing Elec. Document Prod.*, Principle 3 (2007).
13. See, e.g., *Kleen Prods.*, No. 10 C 5711 at \*18; see also *The Sedona Principles*, *supra* note 12, Principle 6 ("Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information.").
14. See, e.g., *EORHB, Inc. v. HOA Holdings LLC*, No. 7409-VCL No. 2013 WL 1960621 (Del. Ch. May 6, 2013) (the court had ordered parties to "show cause" as to why they should not use predictive coding; here, the court allowed plaintiffs to use traditional review, but only because all parties agreed that there was no need for plaintiffs to use predictive coding and that no party would be prejudiced by plaintiffs using traditional review).
15. Even if you are unable to persuade your adversary to use predictive coding for its pre-production review, you should still consider using predictive coding on your end to review large-volume productions for key "hot" documents.
16. See *Moore*, 287 F.R.D. at 183 ("[C]omputer-assisted coding should be used in those cases where it will help 'secure the just, speedy, and inexpensive' . . . determination of cases in our e-discovery world.").
17. "A party that uses advanced analytical software applications and linguistic tools in screening for privilege and work product may be found to have taken 'reasonable steps' to prevent inadvertent disclosure." Fed. R. Evid. 502, Advisory Committee Note; see also Paul W. Grimm et al., *Federal Rule of Evidence 502: Has It Lived Up to Its Potential?*, 17 Rich. J.L. & Tech. 8 at \*36–37 (2011) (footnotes omitted).
18. See, e.g., *Gross*, 256 F.R.D. at 136; *Ruiz-Bueno v. Scott*, 2013 WL 6055402 (S.D. Ohio Nov. 15, 2013) ("[The Rule 26(f)] discussion can and should include cooperative planning, rather than unilateral decision-making . . . [so that] each party is able to exert some measure of control over the e-discovery process, and, in turn, to have some measure of confidence in its results."); *Progressive Cas. Ins. Co. v. Delaney*, No. 2014 WL 3563467 (D. Nev. July 18, 2014); *Bridgestone Ams., Inc. v. IBM*, 2014 WL 4923014 (M.D. Tenn. July 22, 2014).
19. See, e.g., Electronic Discovery Law, *Local Rules Forms, and Guidelines of United States District Courts Addressing E-Discovery Issues*, [www.ediscoverylaw.com/local-rules-forms-and-guidelines-of-united-states-district-courts-addressing-e-discovery-issues](http://www.ediscoverylaw.com/local-rules-forms-and-guidelines-of-united-states-district-courts-addressing-e-discovery-issues).
20. See, e.g., *Gordon v. Kaleida Health*, 2013 WL 2250579 (W.D.N.Y. May 21, 2013); *Hinterberger v. Catholic Health Sys., Inc.*, 2013 WL 2250603 (W.D.N.Y. May 21, 2013).
21. See, e.g., *Moore*, 287 F.R.D. at 186, 193.
22. See, e.g., *In re Actos (Pioglitazone) Prods. Liab. Litig.*, MDL No. 2299 (W.D. La.).
23. See, e.g., *Moore*, 287 F.R.D. 182; *In re Actos*, MDL No. 2299; see also Hon. Craig B. Shaffer, *Defensible By What Standard?*, The Sedona Conference, 2012, at 9 (courts need to know whether "the e-discovery protocol in question produce[d] results that are valid and consistent with the requirements and objective underlying the Federal Rules of Civil Procedure."); Hon. Andrew J. Peck, *Search, Forward*, Law Tech. News, Oct. 2011.
24. Some good examples to consider are the detailed ESI protocols approved in *Moore*, 287 F.R.D. at 192, and *In re Actos*, MDL No. 2299, and the model ESI protocols collected by Electronic Discovery Law, *supra* note 19.