PREDICTIVE CODING:
A TRIAL COURT JUDGE’S PERSPECTIVE

Judge Henry Coke Morgan, Jr.*

INTRODUCTION

Achieving “just, speedy, and inexpensive” conduct in discovery will not succeed without the exercise of integrity and good faith on the part of counsel. Requiring lead counsel to be involved at the earliest stages of the discovery and predictive coding process is crucial. Less experienced counsel are more likely to fall victim to the rationale of using zealous advocacy as a basis for failing to cooperate with opposing counsel in reaching the proper protocols for efficiently producing electronically stored information ("ESI").

Good faith and cooperation will prove more important than technology in solving problems with ESI. The court’s first step should be to encourage counsel to reach a timely agreement on the production of ESI regardless of whether they select the most technologically-advanced protocols. Predictive coding is currently the most advanced form of technologically-assisted research ("TAR") and is the focus of this Article.

However, if counsel fail to reach an agreement or if their agreement fails to produce results, predictive coding appears to be the preferred path to promoting the objectives of Federal Rule of Civil Procedure One ("Rule One").

* Judge Henry Coke Morgan, Jr., was appointed to the United States District Court for the Eastern District of Virginia by President George H.W. Bush in 1992 and is presently serving as a Senior Judge. He is also an adjunct professor at the Regent University School of Law having received his BS and JD from Washington and Lee University in 1957 and 1960, respectively, and a Masters in Judicial Process from the University of Virginia in 1998. He wishes to thank his law clerks Maxwell Thelen and Alisha Burgin for their valued assistance in the preparation of this Article.


3 “These rules govern the procedure in all civil actions and proceedings in the United States district courts, except as stated in Rule 81. They should be construed and administered to secure the just, speedy, and inexpensive determination of every action and proceeding.” Fed. R. Civ. P. 1.
problem.” If predictive coding is the state of the art in electronic discovery, trial courts may adopt it with respect to ESI without waiting for further precedent or rule-making.

This Article suggests that trial courts should adopt predictive coding as it appears to be an improvement upon simple keyword- or manual-search systems. As the cost of civil litigation continues to escalate, primarily driven by ever increasing discovery and pre-trial motion practice, immediate solutions should be sought, or the civil jury trial will face obsolescence as a method of resolving civil disputes over money damages and property rights. Minimizing the costs of handling ESI is only one of the many facets of controlling the costs of civil litigation, but it is an increasingly important facet, and controlling litigation costs has important long-term consequences for both the bench and the bar. This Article, therefore, addresses the question of whether predictive coding promotes the underlying principles of Rule One. Given the threelfold nature of these principles, the analysis proceeds in three parts to reach the conclusion that predictive coding represents a positive step toward achieving just, speedy, and inexpensive trials.

I. IS PREDICTIVE CODING JUST?

The Federal Rules of Civil Procedure provide federal trial judges with an excellent framework within which to encourage the efficient handling of ESI. Beginning in 2006, these rules treated ESI as a separate subset of documents subject to production pursuant to Rules 26, 34, and 45. The rules recognize proportionality in balancing the relevance and importance of ESI with the costs of its production, and provides cost shifting where ESI is inaccessible and sanctions where production efforts fall below acceptable standards. However, these rules

---


5 For a discussion addressing the various costs associated with ESI see Mia Mazza et al., In Pursuit of FRCP 1: Creative Approaches to Cutting and Shifting the Costs of Discovery of Electronically Stored Information, 13 RICH. J.L. & TECH., no. 3, art. 11, Spring 2007, at 3–5, http://law.richmond.edu/jolt/v13i3/article11.pdf.

6 See FED. R. CIV. P. 26(a)(1)(A)(ii)-(iii) (requiring initial production, including damage information) and advisory committee's note to 2006 amendments. The early production of comprehensive damage information is most important as it is often the case in civil litigation that the damages available at the end of the day do not justify the cost of the litigation.

7 See FED. R. CIV. P. 34 and advisory committee's note to 2006 amendments.

8 See FED. R. CIV. P. 45 and advisory committee's note to 2006 amendments.

9 See FED. R. CIV. P. 26(b)(2)(B), (g)(3) and advisory committee's note to 2006 amendments; FED. R. CIV. P. 37(b)–(e).
wisely do not attempt to classify what forms of ESI qualify as accessible or what should be deemed inaccessible due to the difficulty or cost of production.\footnote{The Fourth Circuit, in a case where sanctions were not involved, ruled upon the issue of what forms of ESI discovery may be included as taxable costs. See Country Vintner of N.C., LLC v. E. & J. Gallo Winery, Inc., 718 F.3d 249, 260 (4th Cir. 2013).} Case law, industry principles, and law reviews have categorized ESI for the purpose of discovery and furnish useful guidelines.\footnote{See, e.g., D’Onofrio v. SFX Sports Grp., Inc., 254 F.R.D. 129, 132 (D.D.C. 2008); THE SEDONA CONFERENCE, THE SEDONA PRINCIPLES: BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION 11 (Jonathan M. Redgrave et al. eds., 2d ed. 2007) [hereinafter SEDONA PRINCIPLES], available at https://thesedonaconference.org/publication/ sedona-principles-addressing-electronic-document-production-second-edition; see also Thomas Y. Allman, Local Rules, Standing Orders, and Model Protocols: Where the Rubber Meets the (E-Discovery) Road, 19 RICH. J.L. & TECH., no. 3, art. 8, Spring 2013, http://jolt.richmond.edu/v19i3/article8.pdf.} However, such guidelines are a moving target, and the rapid evolution of technology suggests rule-makers should continue to leave such fine-tuning to the discretion of the trial judge based upon the developing state of the art.

The term “predictive coding” does not lend itself to a single precise definition.\footnote{See Nicholas Barry, Note, Man Versus Machine Review: The Showdown Between Hordes of Discovery Lawyers and a Computer-Utilizing Predictive-Coding Technology, 15 VAND. J. ENT. & TECH. L. 343, 354–55 (2013).} Generally, it may be defined as a five-step process: (1) performance of a form of keyword search, which may include such concepts as clustering, categorizing, culling, and threading through which raw electronic data is organized into a sample set of documents potentially subject to production;\footnote{Id. at 354.} (2) review of the sample by the lead attorneys who will codle the sample documents as responsive or nonresponsive to discovery or as privileged or work product,\footnote{See id.} which is a process known as seeding;\footnote{See Daniel Martin Katz, Quantitative Legal Prediction—or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry, 62 EMORY L.J. 909, 946 (2013).} (3) use of the seeded documents to enable the software to learn what is relevant and subject to discovery;\footnote{Barry, supra note 12, at 354.} (4) organization of the discoverable documents for production, perhaps in terms of percentages of probability for individual documents, and creation of sample subsets of discoverable and non-discoverable documents for use in the verification process;\footnote{See id.} and, finally, (5) manual
review of the sample subsets by lead counsel to verify the accuracy of the process.\textsuperscript{18}

Manual review has always been the norm in handling paper documents and has been adapted to document-by-document screen searching for relevant ESI.\textsuperscript{19} It is interesting to note who has been utilized to perform this search by law firms. A list that is not necessarily in historic order includes: (1) entry-level associates; (2) paralegals (full time or contract); (3) contract lawyers; and (4) third-party providers, including offshore providers. New associates probably did not envision this type of activity when they applied to law school or when they accepted employment with their respective firms, but may be compelled to perform such searches as a rite of passage. Contract lawyers and paralegals likely foresaw what type of work they would be performing, but the issue of oversight is still important here. Although contract lawyers and paralegals may willingly undertake this type of labor, particularly in the current legal market, this does not relieve the fatigue, boredom, and inattention inherent in a document-by-document manual review.\textsuperscript{20} Such manual review is not likely to result in perfection, and its flaws will only be exacerbated as the volume of documents increases.\textsuperscript{21}

The use of third-party providers may involve the same level of personnel through a different entity and produce similar problems. Because third-party providers may have the same level of personnel performing similar tasks as within the law firms, oversight by more experienced lawyers will be necessary.

An analogy may be appropriate here. For most of the twentieth century, home closings were a staple source of practice and significant income for many lawyers. Typically, a title search fee of one percent of the purchase price or loan amount plus some additional fee for document

\textsuperscript{18} See \textit{id.} at 354–55. It is not suggested that substituting lead counsel for another person or entity is a solution. Studies have shown that the participation of lead counsel at the step-two seeding process and step-five verification process promotes efficiency and saves time in the overall process. See Adam M. Acosta, \textit{Predictive Coding: The Beginning of A New E-Discovery Era}, RES GESTAE, Oct. 2012, at 8, 8–9. There are also intangible benefits to bringing lead counsel together early in the litigation. For example, discovery issues tend to be solved more quickly.


preparation or review were charged to the buyer. As home prices escalated, the marketplace would no longer support the one percent fee, and the gold standard of title searches, which utilized practicing lawyers to search grantor indices and judgment liens, gave way to searches by paralegals with reduced closing fees to lawyers. The next evolution involved title searches and closings by title insurers and resultant reductions in participation by lawyers at any level. The marketplace would not continue to support the cost of lawyers performing these ministerial tasks, and lawyers are increasingly being replaced by title companies in the home-closing process.

Similarly, even if lawyers or trained and supervised laypersons were available to manually search ESI, they would likely be challenged by the volume of e-documents while simultaneously being pressured to control the time and cost of the search. As occurred in the home-closing area of law practice, the marketplace will likely doom manual searches of ESI to instances where its volume is large enough to justify what is admittedly a complex process.

The next development after manual searches of ESI has been a keyword search, not unlike the format used in Westlaw and Lexis legal research programs. While this process has been accepted by the courts as an improvement upon manual searches, it has not proven to be an efficient solution. The volume of documents necessary to justify moving from a manual search to predictive coding, however, remains unclear.

---

22 See, e.g., Goldfarb v. Va. State Bar, 421 U.S. 773, 775–76 (1975) (implying that in 1971, all attorneys in Fairfax County, Virginia, were charging this one percent fee).


Some have estimated it to be approximately 100,000 documents.\textsuperscript{28} The question may arise if there is some intermediate volume of documents where a keyword search is more just, speedy, and inexpensive than the direct transition from manual to predictive coding. Since the first step in the predictive-coding process is similar to a keyword search, counsel should be able to make the judgment as to whether the keyword search is adequate at this point. If it is not adequate, the predictive-coding process should proceed.

If handled professionally, and compared with the alternatives of manual and keyword searches, predictive coding should satisfy the reasonableness standard of Rule 26(g).\textsuperscript{29} Professional handling includes the marshaling of all potentially discoverable documents for the keyword search, participation by senior counsel in the seeding and verification processes, and finally, transparency throughout the entire process.

II. IS PREDICTIVE CODING SPEEDY?

How is the trial judge to guide the efficient handling of ESI? It will require early and close attention to establishing protocols for the preservation and prompt production of ESI. In those cases where the court must resolve issues involving ESI, the trial judge must be prepared to rule promptly, or expensive delays and burgeoning discovery costs may result. A valuable tool is a requirement for periodic reports from counsel on the general progress of discovery as well as ESI in particular, which should reveal disputed issues in a timely fashion. If problems arise or if counsel cannot agree at the outset, should the trial judge require the parties to use predictive coding? The court’s initial order covering ESI may prescribe predictive coding as the default provision where counsel are unable to agree upon a protocol. This may cause difficulties where the parties are not familiar with the predictive coding process, and although experienced third-party providers are available, the cost of some could be an issue. Nevertheless, with guidance from the court and experience in usage, predictive coding should solve more problems than it creates.

In most cases the parties should be able to design the predictive coding protocol and reach an agreement about it. It is expected that lead


\textsuperscript{29} See Fed. R. Civ. P. 26(g)(1)(B)(iii) (requiring attorneys or parties to certify that their discovery requests, responses, or objections are “neither unreasonable nor unduly burdensome or expensive, considering the needs of the case, prior discovery in the case, the amount in controversy, and the importance of the issues at stake in the action”).
counsel for all parties will participate in the two-step process of selecting the seed set of documents for coding.\textsuperscript{30} Step three, translating the seed set into a program which will recognize relevant and discoverable documents, requires technical expertise which might be found among the law firms, in-house with clients, or through third-party providers. Wherever this expertise is found, the source of the program must be transparent and acceptable to all parties, and, if they cannot agree, the court must step in. The process should be compatible with each party’s e-document storage protocols as well as compatible or easily convertible to the programs which will be used for discovery and trial. The fourth step is the application of the coded software to a defined set of e-documents in order to extract those which are relevant, and then to separate and index the privileged and work-product documents from those which must be produced. The product of the fourth step is the creation of subsets for production, which are used in the fifth step for verification of the accuracy of the search. While there are several methods of verification, the preferable approach seems to be random sampling of the subsets, with senior counsel coding the samples as accurate or inaccurate, and then running a statistical analysis to determine the error rate—this process is known as “statistical sampling.”\textsuperscript{31}

Ultimately, the speediness of the discovery of ESI will depend upon the selection of the most effective means of conducting the search. For a relatively small volume of documents, a manual search remains most efficient, but as the volume increases, the need for technology increases and should lead, at some level, to the use of predictive coding. There may be an intermediate level at which the less technologically advanced keyword search will be more efficient.\textsuperscript{32} Another advantage of predictive coding is its ability to handle larger volumes of documents, and its potential to become more efficient as the technology continues to improve and counsel becomes more familiar with its utilization.\textsuperscript{33}

\textsuperscript{30} Barry, supra note 12, at 354. If lead counsel delegates this task, such counsel should be given the opportunity to explain this decision to the court if such seeding leads to later problems. For example, delegating could create problems when lead counsel is the author of documents excludable from discovery because of work product or privilege, and, therefore, is better equipped to deal with such issues.

\textsuperscript{31} Id. at 368–69 (discussing five methods of verification and commending statistical sampling because it was “created . . . to check the quality of a large set of goods (here, documents) for which time and cost prohibit individual quality assurance”).


\textsuperscript{33} See id. at 40.
III. 

Is Predictive Coding Inexpensive?

The answer to whether predictive coding is inexpensive depends upon the volume of ESI and the available alternatives. Third-party vendors have evolved, including one which has secured a method patent upon its predictive coding process.34 While the predictive-coding process appears cumbersome at first inspection, there are studies and experience-based, anecdotal evidence that predictive coding is both more accurate and less costly than manual- and simple-keyword search processes.35 In the step-five verification process, a sample set of documents found to be relevant and irrelevant (as well as privileged or unprivileged) are examined by lead attorneys to determine the accuracy of the software. Again, a relatively few hours by lead counsel should prove more effective than a large number by less experienced attorneys or lay providers.36 If the statistical accuracy is unsatisfactory, then the process must be repeated.37 While studies have suggested that predictive coding is already more accurate than a manual search,38 technological


36 Acosta, supra note 18, at 8.

37 Barry, supra note 12, at 354.

developments should continue to enhance the accuracy and cost-effectiveness of predictive coding, and the mind-numbing process of manual review will only become more impractical as the quantity of ESI continues to multiply.39

CONCLUSION

Accordingly, there is a sound basis for moving on from the treatment of manual review as the gold standard.40 Predictive coding is currently the most technologically advanced and accurate system of identifying and producing ESI. However, technology cannot replace integrity and good faith in the discovery process—it can only make the system operate more efficiently when it receives the proper input. The discovery process is at the root of professionalism in trial practice41—it can only operate justly if counsel and parties to the litigation endeavor to produce all discoverable documents and do not tolerate those who do not. Above all, the courts must not tolerate discovery abuse.42 The rules and precedent provide methods for cost shifting and sanctions for careless and deliberate discovery abuse, and courts must be willing to impose them in appropriate cases.43 It is often difficult for courts to determine fault or misconduct, and counsel must be prepared to document such shortcomings in order for courts to perform their duty. Cost shifting, monetary sanctions, and even directed verdicts are not enough for certain levels of discovery abuse—disbarment and criminal prosecution


41 See Judge Jesse G. Reyes, The Role of the Judiciary in Fostering Professionalism and Civility, THE BENCHER, Nov./Dec. 2011, at 9–10 (discussing professionalism and mentioning the need to reliably provide documents).

42 Silvestri v. Gen. Motors Corp., 271 F.3d 583, 590 (4th Cir. 2001) (alteration in original) (quoting United States v. Shafer Equip. Co., 11 F.3d 450, 457 (4th Cir. 1993)) (“The courts must protect the integrity of the judicial process because, ‘[a]s soon as the process falters . . . the people are then justified in abandoning support for the system.’”); see Taylor v. Mitre Corp., No. 1:11-cv-1247, 2012 WL 5473573, at *1–3 (E.D. Va. Nov. 8, 2012) (adopting the magistrate judge’s report and recommendation for dismissal where the plaintiff used a computer program to permanently wipe his hard drive of evidence).

may be appropriate.\textsuperscript{44} Much is at stake. The high cost of civil litigation threatens the civil trial as the gold standard for dispute resolution. Discovery and its abuse are prime drivers of this escalation of costs, and every effort must be made by courts and attorneys to control such costs. While technology has the potential to uncover discovery abuse, the process of uncovering is itself costly, as it will frequently require an expensive search of otherwise inaccessible documents.

Alternative dispute resolution is a valuable aid in resolving civil litigation, but its purpose should not be to replace the civil trial. If judges and juries become obsolete in resolving civil disputes, who is next? May paraprofessionals replace licensed attorneys in appearing before mediators or arbitrators?

Yes, technology can help, and predictive coding should help control costs with increased usage. As technology progresses, predictive coding may be displaced by a more efficient system, but until such time, it is the best available system for ESI production. However, technology will never replace professionalism as the most important factor in the just, speedy, and inexpensive determination of civil actions.

\textsuperscript{44} See United States v. Lundwall, 1 F. Supp. 2d 249, 250 (S.D.N.Y. 1998) (holding that the crime of obstruction of justice under 18 U.S.C. § 1503 covers willful destruction of documents during civil litigation); Peter Vieth, Murray Agrees to a 5-Year Suspension, VA. LAW. WKLY., Aug. 5, 2013, at 2 (reporting a five year suspension from the practice of law for discovery abuse among other things); see also Justin P. Murphy & Matthew A.S. Esoworthy, The ESI Tsunami: A Comprehensive Discussion About Electronically Stored Information in Government Investigations and Criminal Cases, CRIM. JUST., Spring 2012, at 31, 33 (mentioning multiple cases where discovery abuses were referred to the United States Attorney for criminal prosecution).