MEMORANDUM

TO: Commercial Division Advisory Council

FROM: Subcommittee on Procedural Rules to Promote Efficient Case Resolution ("Subcommittee")

DATE: December 11, 2017

RE: Proposal for a Rule Concerning the Use of Technology-Assisted Review in Discovery

---

INTRODUCTION

It is generally agreed that the most expensive stage of complex commercial litigation today is document review. A 2012 RAND study found that document review consumes on average 73% of the total cost of document production in cases involving electronic discovery, notwithstanding such common economies as the use of vendors to do first-level document review.[1] Conducting this resource-intensive stage of litigation in the most efficient manner consistent with defensible results is therefore in the best interest of both litigants and the judicial system. Sophisticated litigants know that the use of technology-assisted review—of which there are many types, ranging from widely used software tools like keyword searching to more sophisticated algorithmic technologies such as predictive coding—can yield substantial cost savings, as well as streamline and accelerate document review and production. The courts of New York State thus would be well-advised to encourage parties to consider the use of technology-assisted review in appropriate cases to speed discovery and reduce its cost.

---

[1] NICHOLAS M. PACE & LAURA ZAKARAS, RAND INST. FOR CIVIL JUSTICE, WHERE THE MONEY GOES: UNDERSTANDING LITIGANT EXPENDITURES FOR PRODUCING ELECTRONIC DISCOVERY, at xv-xvi, 25-27, 41 (2012).

Although technology-assisted review has been available for years, neither the CPLR nor, for that matter, the Federal Rules of Civil Procedure address whether, in what circumstances, or how a party may use technology-assisted review to fulfill its disclosure obligations. In the federal courts, however, the judiciary has provided some guidance through decisions addressing e-discovery issues. In contrast, the New York State courts—including in the Commercial Division, where the costs of document review are likely to be most burdensome—have provided little analogous guidance.

To fill that gap, the Subcommittee proposes a new rule for the Commercial Division addressing technology-assisted review. The proposed rule would do no more than to confirm that technology-assisted review is a legitimate disclosure tool that parties may make use of in appropriate cases, as many are already doing, and that, as with any other document review, the producing party is best situated to determine in the first instance whether and how to use technology-assisted review. The proposed rule would not limit in any way the presiding justice's oversight of the discovery process, nor would it endorse or require any particular kind of technology-assisted review. By supporting the use of technology-assisted review in appropriate cases, however, the proposed rule would make clear that the Commercial Division is receptive to technological innovations that lessen the burdens and cost of complex litigation.

THE PROPOSED RULE

The proposed rule, which might be incorporated as a subpart of current Rule 11-e of the Rules of the Commercial Division, would read as follows:

The parties are encouraged to use the most efficient means to review documents, including electronically stored information ("ESI"), that is consistent with the

parties' disclosure obligations under Article 31 of the CPLR and proportional to the needs of the case. Such means may include technology-assisted review, including predictive coding, in appropriate cases.

DISCUSSION

Background. Litigants in complex commercial cases today use a wide range of technology-assisted review techniques to facilitate the review of what is often an enormous volume of electronically stored information ("ESI"). In such document-intensive cases, human review of each and every collected document for responsiveness can be slower, more costly, and less accurate than the appropriate use of technology-assisted review,[2] which relies on software to help identify potentially irrelevant documents for culling from a large data set, to group together similar documents so as to promote efficient review and consistency of results, or to "teach" a computer to recognize those documents that are most likely to be responsive.

The threshold challenge faced in reviewing a large volume of ESI is that most ESI is unstructured, meaning that it is not organized in any predetermined way. The most common example of unstructured data in the disclosure context is email, which has few predetermined data fields and typically is stored without regard to subject matter. Review of ESI thus often begins by collecting a large volume of unstructured ESI, frequently limited only by custodian and date range, and then running a *keyword search*, which uses software to identify words or phrases that are likely to be found in responsive

[2] *See, e.g., id.* at 55-58, 61-69.

documents, to identify the documents to be reviewed.[3] A more sophisticated variant is *concept searching*, which uses advanced technology to identify documents incorporating concepts similar to the specific search terms used.[4]

The efficiency of the ensuing review and consistency of results can be enhanced through techniques to group similar or related documents together, such as *email threading*, which packages together email strings and any attachments as one chronological thread;[5] *near-duplicate identification*, which groups together similar documents based on their textual similarities (*e.g.*, different drafts of a document);[6] and *clustering*, which uses conceptual analytics technology to group and categorize similar documents.[7]

While these common techniques can help to cull a data set and organize it for review, none of them obviates the need for human review for responsiveness. The form of technology-assisted review generally referred to as *predictive coding* purports, however, to do just that. Predictive coding uses a "machine learning algorithm to distinguish relevant from non-relevant documents, based on subject matter experts'

---

[3] *See, e.g.*, THE SEDONA CONFERENCE, COMMENTARY ON DEFENSE OF PROCESS: PRINCIPLES AND GUIDELINES FOR DEVELOPING AND IMPLEMENTING A SOUND E-DISCOVERY PROCESS 25 (Public Comment Version, 2016).

[4] *See, e.g.*, *Concept Searching*, RELATIVITY.COM, https://www.relativity.com/relativity/Portals/0/Documents/8.0%20Documentation%20Help%20Site/Content/Features/Analytics/Concept%20searching.htm (last visited Nov. 29, 2017).

[5] *See* Nik Balepur, *5 Email Threading Facts That May Surprise You*, THE RELATIVITY BLOG (Apr. 16, 2015), http://blog.kcura.com/relativity/blog/5-email-threading-facts-that-may-surprise-you.

[6] D4, *Near-Duplicate Detection Finds Documents No One Thought Could be Found*, D4 CASE STUDIES BLOG (June 11, 2015), http://d4discovery.com/discover-more/near-duplicate-detection-finds-documents-no-one-thought-could-be-found#sthash.tI5DevpH.dpbs; EQUIVIO, CHOOSING A NEAR-DUPLICATE IDENTIFICATION SOLUTION (2012), http://www.equivio.com/files/files/White%20Paper%20-%20Choosing%20A%20Near-Duplicate%20Identification%20Solution.pdf.

[7] *Document Clustering for eDiscovery Review*, CLOUDNINE, https://www.ediscovery.co/legacy/document-clustering/(last visited Nov. 29, 2017).

coding of a training set of documents."[8] Predictive coding uses computers to extrapolate human judgments about responsiveness, based on human review of a sample "seed set" or "training set" of documents, across the remaining document collection.[9] Because predictive coding requires an upfront investment of time in "teaching" the computer to recognize the characteristics of responsive documents, it generally is cost-effective only when dealing with a large volume of unstructured ESI, but in those circumstances it has the potential to enhance the speed, accuracy, and cost-effectiveness of document review.[10]

Rationale for the Proposed Rule. Both federal and state courts have endorsed the use of technology-assisted review. The United States District Court for the Southern District of New York, for example, has noted that "[p]redictive coding is an automated method that credible sources say has been demonstrated to result in more accurate searches at a fraction of the cost of human reviewers." *Chevron Corp.* v. *Donziger*, 11-CV-0691, 2013 WL 1087236, at *32 n.255 (S.D.N.Y. March 15, 2013). Indeed, the Delaware Chancery Court has actually required parties to use predictive coding. *EORHB, Inc., et al.* v. *HOA Holdings, LLC, et al.*, No. 7409, 2012 WL 4896670 (Del. Ch. Ct. Oct. 15, 2012). Courts have noted in particular the utility of predictive coding for reviewing a large volume of ESI. In the Southern District, for example, Magistrate Judge Andrew Peck has observed that "computer-assisted review is an

---

[8] *The Grossman-Cormack Glossary of Technology-Assisted Review*, 7 FED. COURTS L. REV. 8, 26 (2013) (capitals omitted).

[9] *Id.* at 29, 32-33. Other implementations of predictive coding use a "Continuous Active Learning" model in which the computer "learns" while humans review documents, allowing for the re-classification of documents as the software continuously evolves its "understanding."

[10] JOHN TREDENNICK ET AL., TAR FOR SMART PEOPLE: HOW TECHNOLOGY ASSISTED REVIEW WORKS AND WHY IT MATTERS FOR LEGAL PROFESSIONALS 35-41 (2016).

available tool and should be seriously considered for use in large-data-volume cases." *Moore* v. *Publicis Groupe*, 287 F.R.D. 182, 193 (S.D.N.Y. 2012). Another federal district court has granted a plaintiff's request, over the defendant's objection, to use predictive coding to review approximately 2 million documents for responsiveness. *See Bridgestone Ams., Inc.* v. *Int'l Bus. Machs. Corp.*, No. 3:13-1196, 2014 WL 4923014, at *1 (M.D. Tenn. 2014). Foreign courts have likewise recognized the utility of predictive coding in reviewing large volumes of ESI. *See, e.g., Irish Bank Resolution Corp. Ltd & ors* v. *Quinn &* ors, [2015] IEHC 175 (Ir.); *Brown* v. *BCA Trading Ltd.*, [2016] EWHC 1464 (Ch) (Eng.); *Pyrrho Invs. Ltd.* v. *MWB Prop. Ltd.*, [2016] EWHC 256 (Ch) (Eng.).

The proposed rule would make clear that the Commercial Division is sensitive to the cost of document review in complex commercial cases and is in line with other courts, including other centers of high-stakes commercial litigation such as the Southern District and the Delaware Chancery Court, in supporting the use of technology-assisted review, including predictive coding, in appropriate cases. The proposed rule would not, however, prescribe whether or when any particular form of technology-assisted review may or should be used. These technologies are evolving at a rapid rate, so that any effort to prescribe permissible or impermissible methodologies would quickly become obsolete, and in any event the appropriateness of a given methodology can only be determined in the context of the particular case and the data set to be reviewed. Nothing in the proposed rule is intended to limit the role of the presiding justice in supervising document disclosure, *see* CPLR 3104(a), or to insulate the responding party's production from challenge, *see* CPLR 3124.

Need for Proportionality. Regardless of the method a party uses to review a large collection of ESI for responsiveness, the result will not be perfect. "There simply is no review tool that guarantees perfection. . . . [T]here are risks inherent in any method of reviewing electronic documents." *Moore* v. *Publicis Groupe*, 11-CV-1279, 2012 WL 1446534, at *3 (S.D.N.Y Apr. 26, 2012) (affirming Magistrate Judge Peck's acceptance of predictive coding). Courts have recognized that the standard for a review, whether technology-assisted or entirely human, "is not perfection, or using the 'best' tool, but whether the search results are reasonable and proportional." *Hyles* v. *N.Y. City*, 10-CV-3119, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016). "The goal is for the review method to result in higher recall and higher precision than another review method, at a cost proportionate to the 'value' of the case." *Moore*, 287 F.R.D. at 190.

This concept of proportionality is embedded in the Commercial Division Rules. The Preamble to the Rules provides: "The Commercial Division is mindful of the need to conserve client resources, *encourage proportionality in discovery*, promote efficient resolution of matters, and increase respect for the integrity of the judicial process" (emphasis added). Consistent with these principles, the CPLR limits the scope of disclosure to "all matter *material and necessary* in the prosecution or defense of an action." CPLR 3101(a) (emphasis added). Federal procedure is aligned with the CPLR and the Commercial Division Rules in this respect; the Federal Rules of Civil Procedure similarly limit discovery to that which is "proportional to the needs of the case."[11]

[11] Fed. R. Civ. P. 26(b)(1) limits the "scope of discovery" to "any nonprivileged matter that is relevant to any party's claim or defense *and proportional to the needs of the case*, considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of discovery in resolving the issues, *and whether the burden or expense of the proposed discovery outweighs its likely benefit*" (emphases added).

Accordingly, it should not be a legitimate objection to a party's use of predictive coding or other technology-assisted review that the chosen method may not deliver perfect results. If the methodology chosen is reasonable in the circumstances—that is, "if the burden of identifying additional ESI outweighs the need for [additional] discovery and its importance in resolving the issues in dispute"—then it should be deemed sufficient to meet a party's disclosure obligations.[12] To underscore this principle, the proposed rule incorporates proportionality as a relevant consideration in determining the appropriateness of a document review method.

Parties Encouraged to Cooperate. Because the responding party knows best what kinds and volume of documents it has, how they are stored, and what it will cost to review them, "[r]esponding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information." *Hyles* v. *N.Y. City*, 10-CV-3119, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016) (quoting THE SEDONA CONFERENCE, THE SEDONA PRINCIPLES: SECOND EDITION BEST PRACTICES RECOMMENDATIONS & PRINCIPLES FOR ADDRESSING ELECTRONIC DOCUMENT PRODUCTION, at Principle 6 Illustration i (2d ed. 2007), *available at* www.thesedonaconference.org). "Unless [the responding party's] choice is manifestly unreasonable or the requesting party demonstrates that the resulting production is deficient, the court should play no role in dictating the design of the search." *Mortg. Resolution Servicing, Inc.* v. *JPMorgan Chase Bank, N.A.*, 15-CV-0293, 2017 WL 2305398, at *2 (S.D.N.Y. May 18, 2017).

[12] *See id.*

The proposed rule makes clear that, while the responding party is best placed to analyze in the first instance what it believes to be the most efficient means to review its own documents, including ESI, subject to its disclosure obligations under the CPLR, parties are well advised to confer and agree on an appropriate approach to document review, and the proposed rule encourages them to do so. The proposed rule encourages the responding party to consider the most efficient means to meet its obligations, including technology-assisted review where appropriate, but it does not prevent the requesting party from challenging those means as inadequate or a production as incomplete, nor does the proposed rule constrain in any way the presiding justice's oversight of the disclosure process.

CONCLUSION

The proposed rule would simply align the Commercial Division with those courts, state and federal, that have had occasion to consider the appropriate use of technology-assisted review to promote efficiency and proportionality, consistent with the responding party's disclosure obligations. The proposed rule would reserve to the presiding justice, however, the power to determine whether in the circumstances of a particular case a responding party has met its disclosure obligations.