GARBAGE IN, GARBAGE OUT: IS SEED SET DISCLOSURE A NECESSARY CHECK ON TECHNOLOGY-ASSISTED REVIEW AND SHOULD COURTS REQUIRE DISCLOSURE?

Shannon H. Kitzer[*]

TABLE OF CONTENTS

I. INTRODUCTION

"At this point in the evolution of e-discovery, it is probably best for all concerned to 'drop the e,' and recognize that e-discovery encompasses all discovery."[1] Parties to litigation are increasingly faced with vast amounts of

* J.D., University of Illinois College of Law, 2018; B.A., Northwestern University, 2011. This Note is dedicated in memory of my father, Alan J. Kitzer. A special thank you to Eric A. Johnson, and the University of Illinois Journal of Law, Technology & Policy editors, members, and staff for their contributions in editing this Note. Lastly, I extend my sincerest gratitude to my family and friends for their continued love and support—particularly to my mother, Maureen D. Kitzer, and to Fallon M. Kelly, Erin R. Suhajda and Kelly B. Baron.

1. Tracy Greer, *Avoiding E-Discovery Accidents & Responding to Inevitable Emergencies: A Perspective from the Antitrust Division*, ABA SPRING MEETING 1, 9 (Mar. 2017), https://www.justice.gov/atr/page/ file/953381/download.

electronically stored information during the discovery process.[2] To confront the task of sorting through and perhaps reviewing millions of documents, many parties have turned to technology-assisted review (TAR), also referred to as predictive coding.[3] TAR has gained wide acceptance in the e-discovery community,[4] however, now that TAR is being used with greater frequency, parties and courts alike are faced with the challenge of operationalizing its use. Among the most divisive issues with respect to the use of TAR is whether parties should be required to disclose their seed sets—the documents used to train the computer algorithm[5]—as they are a primary check on the validity of the given TAR process a party employs.[6]

At the most basic level, TAR uses computer algorithms to classify and sort documents.[7] With some TAR methodologies, the computer algorithms are developed and the computers are "trained" using seed sets of documents.[8] The seed set documents are traditionally selected through either random or judgmental sampling, although some parties choose to use a combination of the two methods.[9] Attorneys manually review a sample set of documents and code

2. *See* KRISTI DAVIDSON, *A Game of Tug of War: Balancing Broad Discovery Against Burdensome ESI*, *in* NEW DEVELOPMENTS IN EVIDENTIARY LAW IN NEW YORK, 2014 WL 2344837 at *5 (2014) (stating that it is no longer unusual for litigants and attorneys to be faced with terabytes of data that must be preserved, collected, searched and produced).

3. Technology-Assisted Review (TAR) is a "process for prioritizing or coding a collection of Electronically Stored Information using a computerized system that harnesses human judgments of subject matter expert(s) on a smaller set of documents and then extrapolates those judgments to the remaining documents in the collection." *The Sedona Conference Glossary: E-Discovery and Digital Information Management (Fourth Edition)*, 15 SEDONA CONF. J. 305 (2014) [hereinafter Sedona Best Practices] (adopting definition from Maura R. Grossman & Gordon V. Cormack, *The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola, U.S. Magistrate Judge*, 7 FED. CTS. L. REV. 1, 32 (2013)). While the terms "predictive coding" and "computer-assisted review" are often used interchangeably with TAR, this Note will use the term "TAR" unless quoting from a source that uses another term.

4. *See, e.g.*, Hyles v. New York City, No. 10CIV3119ATAJP, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016) (stating "the Court believes that for most cases today, TAR is the best and most efficient search tool.").

5. *See, e.g.*, Shannon Brown, *Peeking Inside the Black Box: A Preliminary Survey of Technology Assisted Review (TAR) and Predictive Coding Algorithms for E-Discovery*, 21 SUFFOLK J. TRIAL & APP. ADVOC. 221, 251 (2016) ("Most TAR algorithms learn, in a computational model sense, from a specified subset of the entire data set. That subset represents a training set, or sometimes called a seed set.").

6. Cases in which this contentious issue has arisen have yielded a variety of outcomes: courts that have encouraged—though not required—disclosure (*see* Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 192 (S.D.N.Y. 2012); Bridgestone Ams., Inc. v. Int'l Bus. Mach. Corp., Case No. 3:13-1196, 2014 WL 4923014, at *1 (M.D. Tenn. July 22, 2014); *In re* Biomet M2a Magnum Hip Implant Prods. Liab. Litig., No. 3:12-MD-2391, 2013 WL 6405156, at *2 (N.D. Ind. Aug. 21, 2013); responding parties disclosing voluntarily (*see* Bridgestone Ams., Inc. v. Int'l Bus. Machines Corp., Case No. 3:13-1196, 2014 WL 4923014, at *11 (M.D. Tenn. July 22, 2014); Fed. Housing Fin. Agency v. JPMorgan Chase & Co., Case No. 1:11-cv-06188-DLC (S.D.N.Y. July 24, 2012) (tr. at 14–15, 24)); courts not requiring disclosure (*see In re* Biomet M2a Magnum Hip Implant Prods. Liab. Litig., Case No. 3:12-MD-2391, 2013 WL 1729682 & 2013 WL 6405156 (N.D. Ind. Apr. 18 & Aug. 21, 2013); *In re* Actos (Pioglitazone) Prods. Liab. Litig., MDL No. 6:11-md-2299, 2012 WL 7861249 (W.D. La. July 27, 2012)); courts requiring disclosure (*see* Indep. Living Ctr. v. Los Angeles, No. 2:12-cv-00551, slip op. at 1–2 (C.D. Cal. June 26, 2014)), and courts citing non-disclosure as a factor for denying the use of TAR entirely (*see* Progressive Cas. Ins. Co. v. Delaney, Case No. 2:11-cv-00678, 2014 WL 3563467 (D. Nev. July 18, 2014)).

7. Brown, *supra* note 5.

8. *Id.*

9. Charles Yablon & Nick Landsman-Roos, *Predictive Coding: Emerging Questions and Concerns*, 64 S.C. L. REV. 633, 639 (2013).

them as either responsive or non-responsive, and then the computer algorithms are developed based on this sample set.[10]

Requesting parties want seed sets to gain some assurance of the reliability and validity of the document productions they are being provided following the use of TAR[11]—TAR is not a "magic wand" and is only as good as the human reviewers creating the model.[12] Errors in seed sets are of particular concern because they are multiplied:[13] "[a] biased seed set coding could lead to large swaths of relevant documents being deemed irrelevant, and a smoking gun could be missed."[14] Because errors in seed sets are necessarily magnified "Garbage In, Garbage Out . . . has become something of a maxim in predictive coding circles . . . simply mean[ing] that if the seed set receives erroneous judgment, the predictions that are generated will multiply the errors."[15] In fact, the concern that the "other side" will not "do a good job with TAR, regardless of how much 'cooperation and transparency' is going on" is a factor that some experienced litigators consider in "angling toward both sides agreeing to attorney review."[16]

However, producing parties are hesitant to provide this information because it exposes their document review process to criticism; it opens the door to parties haggling over how each document in the seed set was coded.[17] Additionally, practitioners and academics alike have argued that seed sets constitute attorney work product and, accordingly, parties cannot be compelled to produce seed sets,[18] though the Courts have yet to definitively make this determination.[19]

---

10. Brown, *supra* note 5.

11. *See, e.g.*, Rio Tinto PLC v. Vale S.A., 306 F.R.D. 125, 127 (S.D.N.Y. 2015) (explaining the respondent is seeking seed sets because it "fears an incomplete response to his discovery").

12. Aaron T. Goodman, *Predictive Coding and Electronically Stored Information Computer Analytics Combat Data Overload*, ARIZ. ATT'Y 1, 26 (July/Aug. 2016) http://www.azattorneymag-digital.com/azattorneymag/20160708?pg=29#pg29.

13. Edward Sohn, *Top Ten Concepts to Understand About Predictive Coding*, ASS'N. CORP. COUNS. (May 22, 2013), http://www.acc.com/legalresources/publications/topten/ttctuapc.cfm?makepdf=1.

14. Richard H. Lowe et al., *Disclosure of Seed Sets: Required to Cooperate or Protected as Attorney Work Product?*, LEGAL INTELLIGENCER (Feb. 18, 2014), http://www.thelegalintelligencer.com/id=1202643336534/Disclosure-of-Seed-Sets-Required-to-Cooperate-or-Protected-as-Attorney-Work-Product?slreturn=20170106002750.

15. Sohn, *supra* note 13.

16. Christine Payne & Michelle Six, *Three Strategic Choices in E-Discovery*, N.Y. L.J. (Feb. 2, 2018, 5:26 PM), https://www.law.com/newyorklawjournal/sites/newyorklawjournal/2018/02/02/three-strategic-choices-in-e-discovery/?slreturn=20180109160442.

17. *See, e.g.*, Hon. Andrew J. Peck, *Introduction to eDiscovery*, *in* EDISCOVERY FOR CORPORATE COUNSEL § 1:6 (Carole Basri & Mary Mack eds., 2018) (explaining how counselors feel "they are educating the adversary or opening the door to contentious responding demands (you suggest 3 key player/custodians and 25 search terms, the adversary demands 10 custodians) . . .").

18. *See, e.g.*, Hon. John M. Facciola & Philip J. Favro, *Safeguarding the Seed Set: Why Seed Set Documents May Be Entitled to Work Product Protection*, 8 FED. CTS. L. REV. 1, 2 (2015) (arguing that "a seed set generated through counsel's exercise of skill, judgment, and reasoning may reflect its perceptions of relevance, litigation tactics, or even its trial strategy" and "that these conclusions regarding key strategic issues—memorialized in counsel's selection of documents—are entitled to protection from disclosure under the attorney work product doctrine since they may reveal counsel's mental processes and legal theories.")

19. *See, e.g.*, Rio Tinto PLC v. Vale S.A., 306 F.R.D. 125, 128 (S.D.N.Y. 2015) (acknowledging absent parties' agreement "the decisions are split and the debate in the discovery literature is robust" without deciding the applicability of the work product doctrine).

First, as background, this Note will briefly introduce the basics of TAR, its role in the e-discovery process, as well as how and why it came to be used in the e-discovery field. Second, this Note will analyze the recent split among courts as to whether parties should be required to disclose seed sets. Whether a party should disclose its seed sets—or be compelled to disclose its seed sets—remains unresolved not just because of novelty, but also because the novelty stretches the application of the doctrines and rules that would ordinarily govern the e-discovery process. Third, this Note will consider whether the Federal Rules of Civil Procedure empower federal courts to require the disclosure of seed sets and will argue that seed sets have a relational or procedural relevance. Fourth, this Note will consider whether seed sets, even if relevant, are nevertheless protected as attorney work product. Fifth, this Note contemplates the potential applicability of the Federal Rules of Evidence and *Daubert* to TAR. Finally, this Note recommends to parties who seek to mitigate discovery disputes that they should reach an agreement (pertaining to their TAR process generally, but specifically to the disclosure of seed sets or alternative methods of validation, if any) in their electronically stored information (ESI) protocols prior to utilizing TAR, or where appropriate, to consider TAR methodologies that do not rely on (or that minimize) the use of seed sets.[20] This Note also recommends to courts that the parties' agreement as to seed sets should be a precursor to approving the use of TAR in a matter and recommends an increased *Daubert*-like role for magistrate judges in assessing the validity of a party's TAR process in the event of disputes.

II. BACKGROUND

The Federal Rules of Civil Procedure were adopted in 1938 to provide for "the just, speedy, and inexpensive determination of every action."[21] Perhaps one of the most impactful factors on the "inexpensive determination" of a cause of action and the expense of litigation has been the increase in electronically stored information.[22]

As electronic processing capabilities and storage became more readily available and affordable, the volume of electronically stored information grew exponentially.[23] Now communication is regularly conducted in a flurry of

20. Hyles v. New York City, No. 10CIV3119ATAJP, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016) (holding that continuous active learning methodology "eliminates issues about the seed set and stabilizing the TAR tool.").

21. FED. R. CIV. P. 1 (1938).

22. *See* FED. R. CIV. P. 26(b) advisory committee's note to 1993 amendment (noting the tremendous increase in the amount of potentially discoverable information caused by the "information explosion of recent decades" and the corresponding increase in discovery costs); *see also* Zubulake v. UBS Warburg LLC, 217 F.R.D. 309, 311 (explaining "[a]s individuals and corporations increasingly do business electronically . . . the universe of discoverable material has expanded exponentially. The more information there is to discover, the more expensive it is to discover all the relevant information until, in the end, 'discovery is not just about uncovering the truth, but also about how much of the truth the parties can afford to disinter'").

23. *See* Davidson, *supra* note 2 (citing Chevron Corp. v Donziger, 2013 WL 1087236 at *33 (S.D.N.Y. 2013) (stating that "[i]ndeed, there is no persuasive evidence that the compliance costs [over $1 million to review 300,000–800,000 documents] are out of line with what would be typical for nonparty witness in complex commercial litigation"); *see also* Sedona Best Practices, *supra* note 3, at 192 n.2 (noting that "[o]ne gigabyte of

emails with many recipients.[24] Additionally, as electronic storage became more affordable and attorneys and their clients were no longer constrained by the number of boxes that fit on their shelves, preservation expectations rose.[25] In turn, the costs associated with reviewing all the data (that could now be stored at a lower cost), increased.[26]

In response to discovery requests, attorneys have a duty to produce "discovery regarding any non-privileged matter that is relevant to any party's claim or defense and proportional to the needs of the case."[27] This includes ESI.[28] Citing the sheer volume of ESI, the Rules committee issued a proportionality requirement in efforts to reduce the expense and burden of electronic discovery.[29] Notably, amendments to the Federal Rules of Civil Procedure with respect to discovery obligations have carved out a large role for magistrate judges: "magistrate judges became prolific with orders and opinions to flesh out the gaps in the rules."[30]

Despite the proportionality requirement, the fact remains that parties have—and will continue to have—high volumes of documents to review. Faced with the increased expense of litigation, parties began to explore technology-assisted review.[31] Generally, TAR refers to "the computer algorithms that classify, or sort, documents into discrete categories."[32]

Prior to the use of TAR, attorneys manually reviewed documents and determined on a document-by-document basis whether a given document was responsive to a discovery request.[33] Manual review is a linear process and as a

---

electronic information can generate approximately 70,000–80,000 of text pages, or thirty-five to forty banker's boxes of documents (at 2,000 pages per box). Thus, a 100-gigabyte storage device (*e.g.*, a personal computer hard drive), theoretically, could hold as much as the equivalent of 3,500 to 4,000 banker boxes of documents . . . Even if only [ten] percent of a computer's available capacity today contains useful or 'usable' information (as distinguished from application programs, operating systems, utilities, etc.).").

24. *See, e.g.*, Peck, *supra* note 17 (noting "email still remains the principle source of ESI").

25. John M. Facciola, *The Grossman-Cormack Glossary of Technology-Assisted Review with Foreword by John M. Facciola, U.S. Magistrate Judge*, 7 FED. CTS. L. REV. 1 (2013) ("Proportionality has been interpreted in the case law to apply to preservation as well as production.").

26. Sedona Best Practices, *supra* note 3, at 192 (noting the decreasing cost of storing larger amounts of data, but the rising cost of reviewing that same data).

27. FED. R. CIV. P. 26(b)(1).

28. *Id.* ("all documents, electronically stored information"); *see also* FED. R. CIV. P 26(b) advisory committee's note to 2006 amendment ("[R]ecognizing that a party must disclose electronically stored information as well as documents that it may use to support its claims or defenses.").

29. *See* FED. R. CIV. P 26(b)(2) (urging consideration of whether "the burden or expense of the proposed discovery outweighs its likely benefit, taking into account the needs of the case, the amount in controversy, the parties' resources, the importance of the issues at stake in the litigation, and the importance of the proposed discovery in resolving the issues.").

30. *See, e.g.*, Carole Basri & Mary Mack, *Introduction*, *in* EDISCOVERY FOR CORPORATE COUNSEL § 1.1 (Carole Basri & Mary Mack eds., 2018) (reviewing decision issued by magistrate judges analyzing discovery obligations).

31. *See, e.g.*, Thomas C. Gricks III & Robert J. Ambrogi, A Brief History of Technology Assisted Review, (Nov. 17, 2015), L. & TECH. TODAY, https://judicialstudies.duke.edu/wp-content/uploads/2017/07/Panel-1-A-Brief-History-of-Technology-Assisted-Review.pdf (discussing findings that "[t]echnology-assisted review can (and does) yield more accurate results than exhaustive manual review, with much lower effort" as "big news" for "lawyers and clients facing spiraling e-discovery costs.").

32. Brown, *supra* note 5.

33. Andrew Peck, *Search, Forward*, L. TECH. NEWS 1, 25 (Oct. 1, 2011), https://www.law.com/legaltechnews/almID/1202516530534/Search-Forward/.

result, documents often were not being de-duplicated and were thus subject to being coded differently by different reviewers.[34]

Due to the volume of electronically stored information, manual, linear review as the sole method of document review has been largely eliminated and, in some matters would be "virtually impossible."[35] Although before parties turned to TAR, they were utilizing keyword searches in attempts to cull through ESI.[36] Keyword searching, however, has been criticized on many grounds.[37] One common criticism is described as the "Go Fish" problem because the way "most lawyers engage in keyword searches is . . . the equivalent 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)."[38] Additionally, keyword searching is prone to returning false positives.[39] Judicial decisions have criticized keyword searching.[40] Notably, several studies have shown that keyword searches "only return between 20-24% of relevant documents, even when the attorneys think they have retrieved 75% of the relevant documents."[41] Accordingly, keyword searching—while still employed by some often as part of a larger TAR methodology—did not appear to be the "answer" to reviewing high volumes of ESI.[42]

TAR is using computer algorithms to classify and sort documents.[43] With some TAR methodologies, the computer algorithms are developed and the computers are "trained" using seed set documents.[44] Attorneys manually review a sample set of documents and code them as responsive or non-responsive and the computer algorithms are developed based on this sample set.[45] The computer identifies the properties of documents in the seed sets and predictively applies the reviewer's coding to other documents.[46] After enough iterative training

34. *Id.*

35. *Id.* at 25–26 ("Because the volume of ESI has made full manual review virtually impossible, lawyers have turned to keywords to cull ESI (particularly e-mail) for further (manual) review.").

36. *Id.* at 26.

37. *See, e.g.*, Makowski v. SmithAmundsen LLC, No. 08 C 6912, 2012 WL 1634832, at *1 (N.D. Ill. May 9, 2012) ("[U]nfortunately, even a well-designed and tested keyword search has serious limitations. Chief among them is that such a search necessarily results in false positives (irrelevant documents flagged because they contain a search term) and false negatives (relevant documents not flagged since they do not contain a search term)."); Sedona Best Practices, *supra* note 3, at 201 ("[S]imple keyword searches end up being both over- and under-inclusive in light of the inherent malleability and ambiguity of spoken and written English (as well as other languages).").

38. Peck, *supra* note 33, at 26.

39. *Id.*

40. United States v. O'Keefe, 37 F. Supp. 2d 14, 24 (D.D.C. 2008); Equity Analytics, LLC v. Lundin, 248 F.R.D. 331, 333 (D.D.C. 2008); Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 260, 262 (D. Md. 2008).

41. Benjamin L. S. Ritz, *Will This Dog Hunt?: An Attorney's Guide to Predictive Coding*, 57 S. TEX. L. REV. 345, 348 (2016).

42. *See, e.g.*, Peck, *supra* note 33 discussing common critiques of keyword searches and observing "[e]ven with keyword searching [as a method of approaching high volumes of ESI], lawyers have turned to certain computer-assisted approaches to further reduce review cost).

43. Brown, *supra* note 5.

44. *Id.*

45. *Id.*

46. Peck, *supra* note 33, at 29.

rounds, when the system's predictions and the attorneys' coding sufficiently coincide (it normally requires the manual review of a few thousand documents), the algorithm is applied to the entire population of documents.[47] Thereafter, counsel normally engages in quality control sampling.[48]

TAR has many advantages over manual linear review, though one major advantage of TAR is its cost effectiveness—on average TAR yields a fifty-fold reduction in the number of documents requiring manual review.[49] Furthermore, some studies found that TAR "can (and does) yield more accurate results than exhaustive manual review, with much lower effort."[50] Additionally, TAR, unlike keyword searching, "can also search through metadata to build a more complete picture of relevance."[51] This ability to simultaneously consider all aspects and data pertaining to a particular document is one aspect that makes TAR "more efficient and cheaper than manual document review."[52]

While parties were beginning to recognize the virtues of TAR, they were hesitant to make the shift without court approval. *Da Silva Moore v. Publicis Groupe* is significant because it was the first judicial decision to "approve" the use of TAR.[53] The Honorable Judge Andrew Peck—who "has issued many leading e-discovery opinions"[54]—advised: "[w]hat 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."[55]

In "approving" the use of TAR, the *Da Silva Moore* court cited five factors, which have tended to reoccur in and drive the analysis of subsequent decisions:

> (i) [T]he parties' agreement to use predictive coding; (ii) the volume of ESI that had to be reviewed (more than three million documents); (iii) the superiority of computer-assisted review to other available alternatives, i.e., linear manual review or keyword searches; (iv) the need for cost effectiveness and proportionality under Fed. R. Civ. P. 26(b)(2)(C); and (v) the transparent process proposed by the defendant.[56]

47. *Id.*

48. *Id.*

49. Maura R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, 17 RICH. J.L. & TECH. 11, 44 (2011).

50. *Id.* at 48; *see also* Herbert L. Roitblat, et al., *Document Categorization in Legal Electronic Discovery: Computer Classification v. Manual Review*, 61 J. AM. SOC'Y INFO. SCI. & TECH. 70, 79 (2010) ("On every measure, the performance of the two computer systems was at least as accurate (measured against the original review) as that of a human re-review.").

51. Ritz, *supra* note 41.

52. *Id.*

53. Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 182, 187, 192–93 (S.D.N.Y. 2012).

54. *Hon. Andrew J. Peck*, SEDONA CONF., https://thesedonaconference.org/bio/peck-andrew (last visited Mar. 23, 2018).

55. *Da Silva Moore*, 287 F.R.D. at 182, 187, 192–93.

56. *Id.* at 192.

However, not all TAR methodologies are the same[57]—they are "not all created equally; and selecting the right one is critical (think screw driver v. hammer here). Broadly, there are tools driven completely by algorithms and there are tools that incorporate linguistic modeling."[58] Some of the main differences hinge on the training process with respect to the algorithm. Three main types are: continuous active learning, simple active learning, and simple passive learning.[59]

Continuous active learning (CAL) begins with the creation of a seed set that is used to train a learning algorithm.[60] Simple Active learning (SAL) also begins with the creation of a seed set that is used to train a learning algorithm,[61] however, the seed set is selected using keywords, a random section, or a combination of both methods.[62] Essentially, the human operators start "with whatever relevant documents [they] can find, often through keyword search, to initiate the training" and "[f]rom there, the computer presents additional documents designed to help train the algorithm."[63]

With SAL, and unlike CAL, the subsequent training documents to be reviewed and coded are selected using uncertainty sampling, a method that selects the documents about which the learning algorithm is least certain—"in effect, the machine learning algorithm is trying to figure out where to draw the line between the two based on the documents in the control set you created to start the process."[64] These documents are added to the training set, and the process continues until the benefit of adding more training documents to the training set would be outweighed by the cost of reviewing and coding them (a point often referred to as "stabilization").[65] Simple passive learning (SPL), unlike CAL or SAL, generally relies on the random selection of documents, and not the learning algorithm, to identify the training set.[66] So-called "passive learning" because "the machine-learning algorithm plays no role in the selection of training documents."[67]

One key difference to note: "[i]f the TAR methodology uses 'continuous active learning' (CAL)" (as opposed to simple passive learning (SPL) or simple

57. Sharon D. Nelson & John W. Simek, *Predictive Coding: A Rose by Any Other Name*, 38 A.B.A. L. PRAC. 4, 20 (July 2012), https://www.americanbar.org/publications/law_practice_magazine/2012/july-august/hot-buttons.html.

58. Payne & Six, *supra* note 16.

59. *See* Gordon V. Cormack & Maura R. Grossman, *Evaluation of Machine-Learning Protocols for Technology-Assisted Review in Electronic Discovery*, PROC. 37TH INT'L ACM SIGIR CONF. RES. & DEV. INFO. RETRIEVAL 153, 154 (describing and comparing continuous active learning, simple active learning, and simple passive learning) [hereinafter Cormack & Grossman].

60. *Id.* at 154.

61. *Id.*

62. *Id.*

63. John Tredennick, *Continuous Active Learning for Technology-Assisted Review (How It Works and Why It Matters for E-Discovery)*, CATALYST (Aug. 13, 2014), https://catalystsecure.com/blog/2014/08/continuous-active-learning-for-technology-assisted-review-how-it-works-and-why-it-matters-for-e-discovery/.

64. *Id.*

65. Cormack & Grossman, *supra* note 59, at 154.

66. *Id.*

67. Trendennick, *supra* note 63.

active learning (SAL)), "the contents of the seed set is much less significant" because:[68]

> the documents leveraged for the initial iteration should ultimately hold no more significance than any others in the production set. With continuous learning, the predictive coding engine doesn't "stabilize" after a few rounds, thus ending the education of the algorithm prematurely. Rather, the iterations continue until concluded at counsel's discretion. With each iteration, documents coded as relevant are fed into the system to refine the training, which repeatedly updates the rankings of all documents in the database accordingly. At the end of the day, the entire production set has essentially been leveraged for training. Short of including documents withheld for privilege and the limited number of irrelevant documents reviewed before the first iteration, the producing party has already shared the "seed set" upon delivering a production.[69]

Recognizing both the cost-saving benefits and the potential for results more accurate than manual review, many courts have encouraged parties to use TAR.[70] Importantly, courts have also pointed to the availability of TAR in rejecting arguments that certain discovery requests were burdensome.[71] Some courts have even taken the next step and ordered parties to consider using TAR,[72] or awarded attorneys' fees for TAR-related costs.[73] However, as parties embraced TAR, courts began to confront how this technology should be operationalized in the e-discovery process and what should be required of parties to use the technology.[74]

---

68. *Rio Tinto PLC*, 306 F.R.D. at 128; *see generally* Cormack & Grossman, *supra* note 59, at 153–62 (evaluating continuous active learning and simple active learning protocols).

69. Hal Marcus, *Ending the Debate on TAR Seed Sets*, OPENTEXT (Mar. 19, 2015), https://blogs.opentext.com/ending-debate-tar-seed-sets/.

70. Nat'l Day Laborer Org. Network v. U.S. Immigration & Customs Enf't Agency, 877 F. Supp. 2d 87, 109 (S.D.N.Y. 2012); *In re* Domestic Drywall Antitrust Litig., 300 F.R.D. 228, 233 (E.D. Pa. 2014); Malone v. Kantner Ingredients, Inc., No. 4:12-CV-3190, 2015 WL 1470334, at *3 n.7 (D. Neb. Mar. 31, 2015) ("Predictive coding is now promoted (and gaining acceptance) as not only a more efficient and cost effective method of ESI review, but a more accurate one.").

71. Chevron Corp. v. Donzinger, No. 11 Civ. 0691, 2013 WL 1087236, at n.255 (S.D.N.Y. Mar. 15, 2013) (pointing to the availability of TAR in rejecting a burden argument and observing that "predictive 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.").

72. *See, e.g.*, Aurora Coop. Elevator Co. v. Aventine Renewable Energy, No. 4:12-civ-230, slip op. at 1–2 (D. Neb. Mar. 10, 2014) (ordering the parties to "consult with a computer forensic expert to create search protocols, including predictive coding as needed, for a computerized review of the parties' electronic records.").

73. *See, e.g.*, Gabriel Techs. Corp. v. Qualcomm Inc., Case No. 09-cv-1992, 2013 WL 410103, at *10 (S.D. Cal. Feb. 1, 2013) (awarding more than $2.8 million in fees incurred for the or the use of "computer assisted, algorithm-driven document review" for "almost [twelve] million documents.").

74. *See, e.g.*, Jason Krause, *Courts and Judges Embrace Predictive Coding, but is it Really Fixing Discovery?*, ABA J. (Apr. 5, 2018), http://www.abajournal.com/magazine/article/courts_and_judges_embrace_predictive_coding_but_is_it_really_fixing_discove/ (commenting that "despite the increased use of predictive coding in litigation, it can be an ill-defined and nebulous process" and quoting Maura R. Grossman as stating that "[t]he technology is powerful, but there is no standardized process or checklist you can follow[.]").

III. ANALYSIS

A. Courts Have Split on Whether Seed Sets Should be Disclosed, in Part Because TAR Does Not "Fit" Within Traditional Applications of Federal Rules

While courts have yet to require a non-consenting party to use TAR[75] absent extenuating circumstances,[76] TAR has now gained wide acceptance.[77] Given the increasing acceptance and use of TAR, the legal discussion has shifted from whether parties may use TAR,[78] to debates concerning how TAR should be used and what should be required of parties who seek to use TAR. Among the yet-determined issues is whether parties should be required to disclose their seed sets. Courts that have addressed this issue have reached different conclusions: some courts have encouraged—though not required—disclosure,[79] some courts have not required disclosure,[80] some courts have required disclosure,[81] and some courts have cited citing non-disclosure as a factor for denying the use of TAR entirely.[82]

Of the courts that have considered whether parties must disclose seed sets, the most frequent outcome has been for courts to encourage—though not require—disclosure.[83] The analysis in many of these cases emphasizes themes of cooperation and transparency.[84] Transparency, by way of disclosing seed sets, provides comfort for both the court and the parties. This emphasis on transparency and cooperation by the courts seems, at least in part, to be motivated by the respective courts' inability to find a firm source of authority to

75. *See* Hyles v. New York City, 2016 WL 4077114, at *1 ("The key issue is whether, at plaintiff Hyles' request, the defendant City (i.e., the responding party) can be forced to use TAR (technology assisted review, aka predictive coding) when the City prefers to use keyword searching. The short answer is a decisive 'NO.'").

76. *See, e.g.*, Indep. Living Ctr. v. Los Angeles, No. 2:12-cv-00551, slip op. at 1–2 (C.D. Cal. June 26, 2014) (ordering, on consent, the use of TAR to be applied to two million documents after "little or no discovery had been completed" before the discovery cutoff and the parties had yet to reach an agreement after "months of haggling.").

77. *See, e.g.*, *Hyles*, 2016 WL 4077114, at *3 ("[T]he Court believes that for most cases today, TAR is the best and most efficient search tool.").

78. *See, e.g.*, Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 182–83 (S.D.N.Y. 2012) (explaining that at the time "[t]o [the court's] knowledge, no reported case (federal or state) has ruled on the use of computer-assisted coding. While anecdotally it appears that some lawyers are using predictive coding technology, it also appears that many lawyers (and their clients) are waiting for a judicial decision approving of computer-assisted review.").

79. *Id.* at 192; Bridgestone Ams., Inc. v. Int'l Bus. Machs. Corp., Case No. 3:13-1196, 2014 WL 4923014, at *1 (M.D. Tenn. July 22, 2014); *In re* Biomet M2a Magnum Hip Implant Prods. Liab. Litig., No. 3:12-MD-2391, 2013 WL 6405156, at *2 (N.D. Ind. Aug. 21, 2013).

80. *Biomet*, 2013 WL 6405156; *In re* Biomet M2a Magnum Hip Implant Prods. Liab. Litig., Case No. 3:12-MD-2391, 2013 WL 1729682 (N.D. Ind. Apr. 18, 2013); *In re* Actos (Pioglitazone) Prods. Liab. Litig., MDL No. 6:11-md-2299, 2012 WL 7861249 (W.D. La. July 27, 2012).

81. *Indep. Living Ctr.*, slip op. at 1–2 (C.D. Cal. June 26, 2014).

82. Progressive Cas. Ins. Co. v. Delaney, No. 2:11-cv-00678, 2014 WL 3563467 (D. Nev. July 18, 2014).

83. *Da Silva Moore*, 287 F.R.D. at 192; *Bridgestone Ams., Inc.*, 2014 WL 4923014, at *1; *Biomet*, 2013 WL 6405156, at *2.

84. *See, e.g., Da Silva Moore*, 287 F.R.D. at 191 ("Electronic discovery requires cooperation between opposing counsel and transparency in all aspects of preservation and production of ESI." (quoting William A. Gross Constr. Assocs., Inc. v. Am. Mfrs. Mut. Ins. Co., 256 F.R.D. 134, 136 (S.D.N.Y. 2009))).

do anything other than encourage the parties to do the "right" thing.[85] Though notably, litigators living in this world of transparency have remarked "[i]f you're working with opposing counsel who is new to things and/or prone to gamesmanship, then 'cooperation and transparency' can feel a lot more like 'drawn and quartered.' Your client will agree, once she sees the bill."[86]

In *Da Silva Moore*, the defendant voluntarily agreed to produce both the documents in the seed set, as well as counsel's coding of these documents.[87] The *Da Silva Moore* court approvingly discussed this level of transparency,[88] as "such transparency allows the opposing counsel (and the Court) to be more comfortable with computer-assisted review, reducing fears about the so-called 'black box' of the technology."[89] Because the disclosure in *Da Silva Moore* was voluntary, the Court did not have to confront the questions of whether the seed sets were discoverable or if it could compel disclosure.[90] However, the TAR process of which Judge Peck approved seems to suggest the discoverability of seed sets as it contemplates the parties conferring multiple times as to the documents in the seed set and the corresponding coding.[91] Importantly, the *Da Silva Moore* court is not alone in its implication of discoverability.[92] Other courts have endorsed plans that featured conferring and collaboration on seed sets by adversarial parties.[93]

As in *Da Silva Moore*, the court in *Federal Housing Finance Agency v. J.P. Morgan Chase & Co.* opined that the voluntary disclosure of training sets was part of the type of cooperation and transparency necessary for "the TAR process to work."[94] However, while courts have expressed a preference for transparency by disclosure of seed sets, one court did acknowledge it is not the only way to demonstrate reliability of the TAR process in a given matter.[95] The court in *Rio Tinto PLC v. Vale S.A.* further acknowledged that transparency—

85. *See, e.g.*, *Biomet*, 2013 WL 6405156, at * 4 ("The Steering Committee knows of the existence and location of each discoverable document Biomet used in the seed set: those documents have been disclosed to the Steering Committee. The Steering Committee wants to know, not whether a document exists or where it is, but rather how Biomet used certain documents before disclosing them. Rule 26(b)(1) doesn't make such information disclosable.").

86. Payne & Six, *supra* note 16.

87. *Da Silva Moore*, 287 F.R.D. at 192.

88. *See id.* ("This Court highly recommends that counsel in future cases be willing to at least discuss, if not agree to, such transparency in the computer-assisted review process.").

89. *Id.*

90. *Id.* at 186–88.

91. *Id.* at 186–88, 200–03; *see also* H. Christopher Boehning & Daniel Toal, *'Seed Set' Documents Should Not Be Discoverable*, 251 N.Y. L.J., at 2 (Feb. 4, 2014), https://www.paulweiss.com/media/2383003/14feb14e-disc.pdf ("[A]n iterative process in which the parties would confer multiple times upon the documents in the seed set . . . suggested that the seed set used in predictive coding is discoverable to an adversary.").

92. Boehning & Toal, *supra* note 91, at 2.

93. *See, e.g.*, *In re* Actos (Pioglitazone) Prods. Liab. Litig., 2012 WL 7861249, at *3–8 (W.D. La. July 27, 2012) (requiring transparency and plaintiff's access to the seed set's responsive and non-responsive documents, but not privileged documents); *see also*, Boehning & Toal, *supra* note 91, at 2 (describing how, in "a pair of decisions from the Western District of New York, Gordon v. Kaleida Health, and Hinteberger v. Catholic Health System, the court appeared to contemplate that the parties would confer as to the documents used in the seed set for a document production based on predictive coding.").

94. Fed. Housing Fin. Agency v. JPMorgan Chase & Co., No. 1:11-cv-06188-DLC (S.D.N.Y. July 24, 2012) (tr. at 9, 14).

95. Rio Tinto PLC v. Vale S.A., 306 F.R.D. 125, 128–29 (S.D.N.Y. 2015).

through disclosure of seed sets—was a way to demonstrate that TAR was used in an effective and reliable way in the matter, and while the court expressed a preference for disclosure of seed sets, it also noted that "requesting parties can insure that training and review was done appropriately by other means, such as statistical estimation of recall at the conclusion of the review as well as by whether there are gaps in the production, and quality control review of samples from the documents categorized as non-responsive."[96] However, it is important to note that "other validation test methods feature a variety of names, costs, and depths of analysis."[97]

When parties do not voluntarily disclose seed sets, courts have been required to decide whether they *must*; and as mentioned, to-date, these cases have yielded a variety of outcomes.[98]

In *In re Biomet M2a Magnum Hip Implant Products Liability Litigation*,[99] the court rejected the plaintiffs' request to compel disclosure of seed sets on the grounds that the documents were not relevant (within the meaning of the Federal Rules of Civil Procedure).[100] Plaintiff argued that it needed access to the entire seed set so that it could evaluate the accuracy of the production and, if necessary, suggest additional search terms.[101] The court, however, rejected their argument and went further to comment it was "puzzled as to the authority behind the Steering Committee's request" given what it perceived to be the clear inapplicability of FRCP 26(b)(1), which makes only relevant, non-privileged information discoverable.[102] In support of its request that the court require defendant to disclose its seed sets, plaintiffs relied on the Sedona Conference Cooperation Proclamation,[103] which "launche[d] a coordinated effort to promote cooperation by all parties to the discovery process"[104] and "has had a significant, salutary, and persuasive impact on federal discovery practice in the age of electronically stored information."[105] Nonetheless, the court denied plaintiffs' request as the Cooperation Proclamation does not "expand[] a federal district

96. *Id.*
97. Marcus, *supra* note 69.
98. *See supra* Section III.B.
99. *Biomet*, 2013 WL 6405156, at *2.
100. *Id.* at *1–2.
101. *Id.*
102. *Id.* ("The Steering Committee knows of the existence and location of each discoverable document Biomet used in the seed set: those documents have been disclosed to the Steering Committee. The Steering Committee wants to know, not whether a document exists or where it is, but rather how Biomet used certain documents before disclosing them. Rule 26(b)(1) doesn't make such information disclosable.").
103. Though not a binding authority, the Sedona Conference is widely regarded as a preeminent authority on all topics related to electronic discovery. *See, e.g.*, John B. v. Goetz, 531 F.3d 448, 459 (6th Cir. 2008) (citing the Sedona Conference's "Sedona Principles" multiple times); Ford Motor Co. v. Edgewood Props., Inc., 257 F.R.D. 418, 424 (D.N.J. 2009) ("The Sedona Principles and Sedona commentaries thereto are the leading authorities on electronic document retrieval and production."); Jay E. Grenig & William C. Gleisner, III, *eDiscovery & Digital Evidence* § 1:7 (2005) ("While not law, the [Sedona P]rinciples are definitely persuasive authority of the first order."). Courts routinely look to the Sedona Conference and its Principles to determine the reasonableness of a proposed discovery process. *See, e.g.*, Victor Stanley, Inc. v. Creative Pipe, Inc., 250 F.R.D. 251, 262 (D. Md. 2008) ("[C]ompliance with the Sedona Conference Best Practices for use of search and information retrieval will go a long way towards convincing the court that the method chosen was reasonable and reliable . . . .").
104. *Cooperation Proclamation*, 10 SEDONA CONF. J. 331 (Fall Supp. 2009).
105. *Biomet*, 2013 WL 6405156, at *2.

court's powers, so they can't provide [the court] with authority to compel discovery of information not made discoverable by the Federal Rules."[106] Notably, however, while the court did not require the defendant to disclose its seed sets, it stated that it found defendant's position "troubling" and "uncooperative," and "urge[d] [defendant] to re-think its refusal [to disclose its seed sets]."[107]

Contrastingly, another court did order (on consent) that a party use TAR, and to avoid subsequent discovery disputes, it ordered that plaintiffs "be involved with and play an active role" in the training process including making relevance determinations in the training documents.[108] Similarly, in *Federal Housing Finance Agency v. HSBC*, the court (in a decision from the bench) required the producing party to give the requesting party full access to the seed set's non-privileged responsive and non-responsive documents.[109]

One court cited non-disclosure of seed sets as a reason for denying the use of TAR.[110] In *Progressive Casualty Insurance Co. v. Delaney*, the court denied plaintiffs TAR proposal (in which they refused to disclose their seed sets), characterizing their unwillingness to disclose as lacking cooperation and transparency, and finding that moving forward with their ESI protocol "[would] only result in more disputes [and] further delay completion of discovery."[111]

B. Seed Sets Have a Relational or Procedural Relevance and Thus Should be Considered Relevant Within the Meaning of the Federal Rules of Civil Procedure

This Note argues that seed sets have a relational or procedural relevance. On their face, the documents comprising seed sets are not relevant within the traditional application of Fed. R. Civ. P. 26(b)(1) as the rule only requires that non-privileged relevant documents are produced, and therefore parties are not required to produce documents within the seed set that have been coded as non-relevant.[112] Documents coded as relevant within seed sets are produced (though not identified as seed set documents) and seed set documents coded as non-relevant are not produced.

However, within the context of the TAR process, even documents within a seed set coded as non-relevant take on a sort of relational relevance because the coding of non-relevant and relevant documents is how the computer is, in

106. *Id.*
107. *Id.*
108. Indep. Living Ctr. v. Los Angeles, No. 2:12-cv-00551, slip op. at 1–2 (C.D. Cal. June 26, 2014).
109. Fed. Hous. Fin. Agency v. HSBC North Am. Holdings Inc., 2014 WL 584300, at *1 (S.D. N.Y. 2014).
110. Progressive Cas. Ins. Co. v. Delaney, No. 2:11-CV-00678-LRH, 2014 WL 3563467, at *11 (D. Nev. July 18, 2014).
111. *Id.*
112. *See Biomet*, 2013 WL 6405156, at *2 (determining that the production of entire seed set of documents was beyond the scope of permissible discovery (per 26(b)(1)) by seeking irrelevant and potentially privileged information).

essence, trained.[113] Non-relevant documents have a relational relevance in a way that documents coded non-relevant per a traditional linear review do not.[114] In this same vein, one practitioner gave the example of the role of a client's document retention policy in the context of traditional discovery in drawing this distinction between substantive and procedural relevance.[115] While the document retention policy is not "*substantively* relevant" to the issues in a given litigation, document retention policies—which are routinely produced in complex litigation matters—are nonetheless "*procedurally* relevant . . . to the document collection and production process, and therefore within the scope of information to be produced pursuant to the Federal Rules of Civil Procedure."[116]

Similar to a document retention policy, seed sets take on a sort of relational or "*procedural*[] relevance," because if the seed set is inaccurate, all data produced will reflect those inaccuracies.[117] Errors are magnified with TAR in a way they are not with linear review, and without accurate training the data TAR yields is substantively worthless, though it still comes at a high price tag.[118] Disclosing seed sets may be necessary to assure the court and opposing party the document review has a defensible process.

C. Even if Seed Sets are Relevant, They Nonetheless may be Protected by the Attorney Work Product Doctrine

To date, much of the judicial analysis of the discoverability of seed sets has turned on relevance,[119] however, practitioners[120] and academics alike have made arguments that seed sets should be protected as attorney work product. Practitioners argue that the judicial emphasis on transparency, cooperation, and smooth discovery and motion practice "fails to recognize that a seed set generated through counsel's exercise of skill, judgment, and reasoning, and therefore may reveal counsel's perceptions of relevance, litigation tactics, or even its trial strategy" and thus "arguably constitute a lawyer's work product, [that] should be protected from discovery since they could reflect its mental processes and legal theories."[121]

As mentioned, the *Biomet* court rejected the plaintiffs' request to compel disclosure of seed sets on the grounds that the documents were not relevant.[122]

---

113. Whitney Street, *Technology Assisted Review: The Disclosure of Training Sets and Related Transparency Issues* (Sept. 30, 2014), https://blockesq.com/wp-content/uploads/2016/06/georgetown-eDiscovery-conference.pdf.

114. *Id.*

115. *Id.*

116. *Id.*

117. *Id.*

118. *Id.*

119. *See supra* Section III.A.

120. Boehning & Toal, *supra* note 91, at 2.

121. Sporck v. Peil, 759 F.2d 312, 318 (3d Cir. 1985) (holding that counsel's selection of certain documents to prepare a client for deposition was protected as opinion work product); Phillip Favro, *The Question of Disclosure in Predictive Coding*, BLOOMBERG L.: BIG L. BUS. (Mar. 19, 2015), https://bol.bna.com/the-question-of-disclosure-in-predictive-coding/ (citing Fed. R. Civ. P. 26(b)(3)).

122. *In re* Biomet M2a Magnum Hip Implant Prod. Liab. Litig., No. 3:12-MD-2391, 2013 WL 6405156, at *1–2 (N.D. Ind. Aug. 21, 2013).

While the *Biomet* court did not address the work product doctrine with respect to seed sets, Defendants argued they should not be required to produce their judgment-selected seed set because it should be protected as work product.[123] Defendants supported their argument, in part, by comparing seed sets to contract attorneys—reasoning that "[t]eaching the predictive coding software to identify relevant documents is indistinguishable from teaching contract attorneys to do the same by using an instruction manual or examples of relevant documents."[124] In response, Plaintiffs likened seed set determinations to keywords or search terms which the Defendants had previously provided without a work product-based objection, and asserted they were "left with no guidance as to what documents Defendants used to train the system, and therefore with no ability to assess whether Biomet's training was adequate in regard to documents Plaintiffs feel are relevant and important to the litigation."[125]

While the *Biomet* court did not rule on the merits of these arguments, these are the types of arguments parties are likely to offer moving forward—so are seed sets more akin to a training manual for a contract attorney or to a set of keywords? Neither is a perfect analogy, but for parties awaiting resolution of whether seed sets are protected by the work product doctrine it may still be instructive to look to how the work product argument fared with respect to keywords.

Though arguments based on the work product doctrine, as laid out in *Hickman v. Taylor*[126] and *Sporck v. Peil*,[127] have been made and assert that "search terms deserve protection from compelled disclosure as opinion work product,"[128] courts have generally rejected the argument.[129] Those who argue that keywords should be a protected work product contend "the process by which [an attorney] select[s] documents to review for litigation . . . 'reveal[s] her legal theories and opinions,' and so the process constitute[s] opinion work product."[130] Relatedly, on a policy basis, those who believe that keywords should be protected as opinion work product fear that if counsel is forced to disclose their keyword searches their search terms will be used against their client to suss out strategy and impressions, and accordingly attorneys will hesitate to search through their data thoroughly in response.[131]

---

123. Defendants' Response to Plaintiffs' Demand for Defendants' Predictive Coding Seed Set, at 3, *In re* Biomet M2a Magnum Hip Implant Prods. Liability Litig., No. 3:12-MD-2391 (N.D. Ind. Aug. 5, 2013) (ECF No. 722).

124. *Id.* at 2.

125. Plaintiffs' Motion for Relief from Defendants' Refusal to Disclose Relevant Documents Used in Predictive Coding at 2, *In re* Biomet M2a Magnum Hip Implant Prods. Liability Litig., No. 3:12-MD-2391 (N.D. Ind. Aug. 5, 2013) (ECF No. 723).

126. Hickman v. Taylor, 329 U.S. 495 (1947).

127. Sporck v. Peil, 759 F.2d 312, 315 (3d Cir. 1985).

128. Sean Grammel, *Protecting Search Terms as Opinion Work Product: Applying the Work Product Doctrine to Electronic Discovery*, 161 U. PA. L. REV. 2063, 2066 (2013).

129. *See, e.g.*, Apple Inc. v. Samsung Elecs. Co. Ltd., No. 12-CV-0630-LHK (PSG), 2013 U.S. Dist. LEXIS 67085, at *40 (N.D. Cal. May 9, 2013) (noting case law suggests seed sets are not protected by work product doctrine).

130. Grammel, *supra* note 128, at 2095 (citing Shelton v. Am. Motors Corp., 805 F.2d 1323, 1326 (8th Cir. 1986)).

131. *Id.* at 2098.

One of the earlier cases to consider the argument[132] resulted in a "flat-out denial of work product protection."[133] The plaintiff sought to compel disclosure of the defendant's search terms "so that it [could] fully evaluate the Defendant's methodology."[134] Over Defendant's work product-based objection, the court found that Defendant had a "duty to demonstrate that its methodology was reasonable," and required disclosure of search terms to establish adequacy.[135] Similarly, other courts reached the conclusion that the disclosure of search terms was required as the search terms were necessary to determine whether an adequate search was conducted.[136] The court in *FormFactor, Inc. v. Micro-Probe, Inc.* elaborated that search terms are "not subject to any work product protection because it goes to the underlying facts of what documents are responsive to Defendants' document request, rather than the thought processes of Plaintiff's counsel."[137] By mid-2013, one court observed that a party who previously "maintained that its search terms and choice of custodians were privileged under the work-product immunity doctrine" had subsequently "abandoned" the argument "no doubt in part because case law suggests otherwise."[138]

An article by the Honorable Judge John M. Facciola, "one of the leading judicial voices for e-discovery,"[139] considers in-depth the question of whether seed sets should be entitled to work product protection and concludes that seeds are work product.[140] However, Judge Facciola acknowledges currently the "[c]ognoscenti and courts disagree whether the identification of seed set documents is work product and entitled to protection from discovery."[141] Nevertheless, Judge Facciola argues that the two key policies underlying *Hickman*, a seminal Supreme Court case articulating the work product doctrine, are equally applicable to seed sets.[142]

The first policy underlying *Hickman* that arguably supports protecting seed sets as work product is "that counsel is entitled to a zone of privacy to prepare its case for trial."[143] Judge Facciola reasons that because documents selected for judgmental seed sets could "reflect the manner in which counsel is pursing discovery . . . the seed set will disclose counsel's thought processes, certain

132. Smith v. Life Inv'rs Ins. Co. of Am., No. 2:07-cv-681, 2009 U.S. Dist. LEXIS 58261 (W.D. Pa. July 9, 2009).

133. Grammel, *supra* note 128, at 2063.

134. *Smith*, 2009 U.S. Dist. LEXIS 58261, at *19.

135. *Id.*

136. *See* Nat'l Day Laborer Org. Network v. U.S. Immigration & Customs Enf't Agency, 877 F. Supp. 2d. 87, 96 (S.D.N.Y. 2012) (requiring production of search terms in order to assess production, but basing the order on the necessity of such disclosure to ensure adequate compliance with the Freedom of Information Act); FormFactor, Inc. v. Micro-Probe, Inc., No. 10-03095, 2012 WL 1575093, at *7 n.4 (N.D. Cal. 2012) (echoing the *Smith* court's reasoning that the adequacy of production could not be assessed without compelling disclosure of search terms).

137. *FormFactor*, 2012 WL 1575093, at *19–20.

138. *Apple Inc.*, 2013 U.S. Dist. LEXIS 67085, at *40.

139. Grammel, *supra* note 128, at 2086.

140. Facciola & Favro, *supra* note 18, at 32

141. *Id.* at 6.

142. *Id.* at 7–8.

143. *Id.* at 7.

conclusions made on the claims and defenses at issue, and/or its strategy for seeking to dispose of the case."[144]

The second *Hickman* policy rationale is that a litigation adversary should not receive "a free ride on the effort and investment of [] counsel in reviewing and selecting documents and in preparing [its claims] or defense[s]."[145] Here, Judge Facciola argues disclosure of judgmental seed sets could provide opposing counsel insight into their adversary's "legal strategy, his intended lines of proof, his evaluation of the strengths and weaknesses of his case" and enable them to prepare accordingly.[146]

Additionally, Judge Facciola argues that the *Sporck* rule supports treating the judgmental seed sets as opinion work product.[147] The United States Court of Appeals for the Third Circuit held in *Sporck v. Peil* that selections of documents could be protected as opinion work product.[148] In *Sporck*, the court held that a lawyer's selection of a "few documents out of thousands" (selected to prepare a deponent) constitutes protected work product.[149] Citing *Hickman*, the *Sporck* court observed that the document selection process was an essential aspect of case preparation[150] and that accordingly "disclosing counsel's document selection would inappropriately reveal these mental processes to a litigation adversary."[151]

Ultimately, Judge Facciola's arguments that seed sets are work product that are similar to those previously advanced with respect to keywords as work product,[152] which courts subsequently rejected as to keywords.[153] Additionally, while arguing that the work product protection should be afforded to seed sets,[154] Judge Facciola acknowledges that at present the Courts seem to disagree and they have yet to extend work product protection to seed sets.[155] If and when a court reaches this question on the merits, the resolution might turn on whether the court views seed sets more akin to a training manual for a contract attorney or to a set of keywords.

144. *Id.* at 8.

145. *Id.* at 7–8 (quoting U.S *ex rel.* Bagley v. TRW, Inc., No. CV94-7755-RAP(AJWx), 1998 U.S. Dist. LEXIS 23585, at *4 (C.D. Cal. 1998)).

146. *Id.* at 8.

147. *Id.* at 32.

148. Sporck v. Peil, 759 F.2d 312, 315 (3d Cir. 1985).

149. *Id.* at 316 (quoting James Julian, Inc. v. Raytheon Co., 93 F.R.D. 138, 144 (D. Del. 1982)); Facciola & Favro, *supra* note 18, at 33.

150. *Sporck*, 759 F.2d at 316–17 (citing Hickman v. Taylor, 329 U.S. 495 (1947)).

151. Facciola & Favro, *supra* note 18, at 24 (citing *Sporck*, 759 F.2d at 317).

152. Grammel, *supra* note 128, at 2066.

153. *See, e.g.*, *Apple Inc.*, 2013 U.S. Dist. LEXIS 67085, at *40 (N.D. Cal. May 9, 2013) (noting case law suggests seed sets are not protected by work product doctrine).

154. Facciola & Favro, *supra* note 18, at 32.

155. *Id.* at 6.

D. In the Event of a Dispute, a Daubert*-style Analysis is Necessary for Trial Courts to Serve Their Gatekeeping Function*

A true *Daubert*-style analysis has yet to be applied to a TAR process,[156] however, a *Daubert*-style analysis would be a logical extension supported by both the gatekeeping rationale of the rule and the judicial system's need.[157] Some courts have hinted that a *Daubert*-style analysis of TAR may be appropriate.[158] Though others have strongly disagreed, pointing to the fact that the Federal Rules of Evidence govern admissibility of evidence at trial as opposed to during the discovery process.[159] While the FRE and *Daubert* have previously not been applied to the discovery and document review process,[160] the discovery process has greatly changed since the days of manual review when an attorney looked at each document and made a relevancy determination.[161] The linear manual review process performed during the discovery phase did not have the potential to affect the admissibility of evidence at trial, the way that TAR does.[162] Additionally, there has been growing support among experts and academics that *Daubert* should apply to TAR.[163]

It is not the content or nature of the quasi-expert opinion by an e-discovery professional that stretches the traditional bounds of FRE 702, but rather the timing.[164] But for the timing, the methods and opinions (as to validity of a given TAR process and the resulting production) of an e-discovery professional are exactly the type of methods and opinions that would otherwise be subjected to FRE 702 and the *Daubert* factors.[165] There is an entire industry of software vendors and e-discovery experts that have cropped up and developed the technologies that automate the TAR process.[166] The software vendors and e-discovery experts—not the attorneys—create the algorithms and can attest to the

156. *See, e.g.*, Da Silva Moore v. Publicis Groupe SA, 2012 WL 1446534 (S.D.N.Y. Apr. 26, 2012) (questioning the potential applicability of *Daubert* analysis).

157. The disclosure of seed sets would factor into a court's *Daubert*-style analysis as they are a key measure of validity of a production or would be a key metric of evaluation for an opposing expert.

158. *See Da Silva Moore*, 2012 WL 1446534, at *2–3 (acknowledging that a Rule 702 analysis may be necessary, however, analysis was "premature . . . under the circumstances of this particular case.").

159. Peck, *supra* note 33, at 29.

160. *See, e.g.*, *Da Silva Moore*, 2012 WL 1446534 at *2–3 (questioning the potential applicability of *Daubert* analysis).

161. *See supra* Section II.

162. Daniel K. Gelb, *The Court as Gatekeeper: Preventing Unreliable Pretrial e-Discovery from Jeopardizing A Reliable Fact-Finding Process*, 83 FORDHAM L. REV. 1287, 1288 (2014).

163. *See, e.g.*, David J. Waxse & Brenda Yoakum-Kriz, *Experts on Computer-Assisted Review: Why Federal Rule of Evidence 702 Should Apply to Their Use*, 52 WASHBURN L.J. 207, 220 (2013) (arguing computer-assisted review should be subject to *Daubert*); *see also* Gelb, *supra* note 162, at 1290 (discussing the logic of employing a *Daubert* and Rule 702-type approach to TAR).

164. See, e.g., Gelb, *supra* note 162, at 1290 ("As technology continues to weave itself into most—if not all—complex litigation, many evidentiary rulings are likely to be impacted by the underlying integrity of the discovery methodologies employed by the parties *before trial*.").

165. *Id.* at 390 ("To argue that one has done what was necessary to produce the evidence is to argue that one's process was scientific and reliable, which is Federal Rule of Evidence 702."); *see also About Us*, H5, https://www.h5.com/about-us/ (last visited Mar. 23, 2018) (describing an e-discovery company's team "of linguists, data scientists and computational analysts.").

166. *See, e.g.*, Waxse & Yoakum-Kriz, *supra* note 163, at 210 (contemplating what implications "software vendors and e-discovery companies [that] have developed more sophisticated technologies to automate and improve the ESI search and review process" will have on the discovery process).

statistical validity of a given search process.[167] Accordingly, TAR is rooted in scientific disciplines including statistical analysis.[168] However, as the "timing" argument goes, *Daubert* and FRE 702 are not applicable because they are "not procedurally ripe at the eDiscovery review phase of litigation . . . unlike at trial, the fact-finder during discovery has yet to be presented with an evidentiary question upon which to make a finding with the assistance of expert testimony."[169]

The timing argument, while in keeping with traditional notions of FRE 702's applicability, fails to accommodate the shift that both ESI and e-discovery have had generally on the litigation timeline[170] and gives parties limited recourse in the face of junk e-discovery science.[171] If discovery that is a product of a TAR process is introduced at trial, the process a party employed "may raise the issue of witness qualification and, therefore, whether the evidence derived from the review is admissible."[172] If a party has produced hundreds of thousands or even millions of documents using a faulty or unreliable method it will accordingly affect the admissibility of substantive evidence at trial.[173] Furthermore, the reliability of the discovery process determines the development of evidence in a case and thus the development of the case itself.[174] Experienced litigators have recognized that "developing e-discovery strategy hand in hand with trial strategy[]" is "a key litigation opportunity" and that "[s]imply put, e-discovery strategy needs to be married to trial strategy from the beginning of the case."[175] Thus, at least in some instances, assessing the reliability of a party's e-discovery process at trial does not suit the parties' interests nor judicial economy, and in some instances, it could be too late.

The motivating "gatekeeper" rationale of FRE 702 and *Daubert* supports their application to TAR.[176] In *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, the U.S. Supreme Court imposed a special obligation on trial judges to "ensure that any and all scientific testimony . . . is not only relevant, but reliable."[177] The Court has since clarified a trial court's gatekeeping function applies to all expert testimony.[178] A trial court's gatekeeping function, in part, requires the trial judge to determine whether an expert's testimony is based "on a reliable foundation."[179] To be based on a "reliable foundation" means an expert's

167. *Id.* at 211.

168. *See* United States v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) (explaining that the effectiveness of searches depends on computer technology, statistics, and linguistics).

169. Gelb, *supra* note 162, at 1288 (citing Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 188–89 (S.D.N.Y. 2012)).

170. *Id.* at 1296 ("Large data discovery review has only increased the need for the court to invoke—when moved—its gatekeeping function earlier in the litigation process.").

171. *Id.* at 1297.

172. *Id.* at 1293.

173. *Id.* at 1288.

174. *Id.* at 1292.

175. Payne & Six, *supra* note 16.

176. *See, e.g.*, Gelb, *supra* note 162, at 1290 ("The gatekeeping function under *Daubert* is particularly pertinent to the issue of whether Rule 702 should apply to litigating the application of TAR methodologies, such as predictive coding, to complex eDiscovery.").

177. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 589 (1993).

178. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 147 (1999).

179. *Daubert*, 509 U.S. at 597.

conclusions are drawn from scientific knowledge and based upon legitimately sound scientific methodology not merely a standard of general acceptance.[180] Because TAR has forged an "inextricable tie between the evidence developed during the *discovery phase* of litigation and the admissibility of that evidence at the trial phase,"[181] in order for a trial court to perform its gatekeeping function, it cannot wait until trial. If the evidence has already come through the "gate" so to speak—*i.e.*, the parties have been relying on document productions in developing their case and preparing for trial—how much good does it do for the gatekeeper to determine the methodology, and thus the evidence, is unreliable after the fact?

Some judges have at least implicitly recognized this timing problem and while they did not conduct a formal *Daubert* hearing, their pre-trial inquiries nonetheless shared the same aim: ensuring reliability.[182] In *Equity Analytics, LLC v. Lundin*, for example, Judge Facciola—noting that determining whether a methodology would be effective "requires knowledge beyond the ken of a lay person (and a lay lawyer) and requires expert testimony that meets the requirements of Rule 702 of the Federal Rules of Evidence"—required a plaintiff to submit an affidavit from its expert explaining why their proposed course of action was technically sound and to explain in detail how the search would be conducted.[183]

Experts are heavily involved in the document review process in a way they never were before. "Litigants must often retain an expert from the [software] service provider to" aid in the implementation of the technology and workflow because of the "differences in offerings of predictive coding technology and the obscurities of their operations."[184] The engagement of e-discovery experts and vendors to perform TAR—who develop algorithms and statistical models to test and verify the document population—necessitates a change in the timing of at least a *Daubert*-style analysis to evaluate the soundness of their methods and the corresponding completeness and reliability of a production in the event of a dispute.

180. *Id.* at 589–93.

181. Gelb, *supra* note 162, at 1288; *see also* United States v. O'Keefe, 537 F. Supp. 2d 14, 24 (D.D.C. 2008) (asserting that utilizing certain TAR methods "is clearly beyond the ken of a layman" and may require an expert opinion).

182. *See, e.g.*, Victor Stanley, Inc. v. Creative Pipe, Inc. 250 F.R.D. 251, 262 (D. Md. 2008) (stating "the party selecting the methodology must be prepared to explain the rationale for the method chosen to the court, demonstrate that it is appropriate for the task, and show that it was properly implemented"); *see also* Equity Analytics, LLC v. Lundin 248 F.R.D. 331, 333 (D.D.C. 2008) (determining whether particular search methodologies "will or will not be effective . . . requires expert testimony that meets the requirements of Rule 702 of the Federal Rules of Evidence."); *see also In re* Actos (Pioglitazone) Prod. Liab. Litig., No. 6:11-MD-2299, 2012 WL 7861249, at *3–12 (W.D. La. July 27, 2012) (engaging in a pre-trial review of the of the search methodology and coding process and instructing parties on how to proceed).

183. *Equity Analytics*, 248 F.R.D. at 333.

184. Charles Vaccaro, *Look Before You Leap into Predictive Coding: An Argument for A Cautious Approach to Utilizing Predictive Coding*, 41 RUTGERS COMPUT. & TECH. L.J. 298 (2015).

IV. RECOMMENDATIONS

Finally, this Note recommends that parties address the disclosure of seed sets in their ESI protocols and that, if necessary, courts be prepared to engage in a *Daubert*-like analysis to ensure parties are appropriately using TAR. It is, however, important to emphasize that litigators should be seeking to keep both the judge and their clients happy by minimizing expensive and time-intensive discovery disputes—"[d]iscovery disputes are most easily resolved at the beginning of a production and are most intractable after production has begun."[185]

To parties who seek to mitigate discovery disputes, this Note recommends that they should reach an agreement (pertaining to disclosure of seed sets) in their ESI protocols prior to utilizing TAR. Ultimately, given the current uncertainty as to both whether seed sets are relevant and whether seed sets are protected as attorney work product, it behooves parties to control their own fate and reach agreement *ex ante*. Voluntary disclosure also provides parties some assurances that they are receiving accurate and complete discovery responses.[186]

To be prepared to meaningfully participate in negotiation of an ESI protocol, litigators should directly engage with TAR and gain a greater understanding of the nuts and bolts of the technology and process. TAR is often referred to as a "black box"[187] but increased familiarity with TAR should yield increased comfort. The more comfortable the attorneys are with the technology, the more able they will be to engage in meaningful discussions about their ESI protocols. Flimsy ESI protocols merely punt the issues down field. However, attorneys who have a solid working understanding of TAR will be able to get to the root of the issues more quickly, will understand the implications of the concessions they are making, and will have a better sense of which disputes are or are not worth having.

To attorneys searching for guidance, they may consider turning to those who have forged ahead. The Department of Justice Antitrust Division is sophisticated and experienced in matters of e-discovery because "[g]iven the size and complexity of the Division's usual investigations" and their "unprecedented number of cases" in recent years, their "experience with electronic discovery cover[s] a wide range of issues."[188] As a result "[t]he Division has developed "robust production specifications" and has incorporated "TAR requirements into Instruction 4 of [their December 2016] Model requests" which parties may wish to consult for guidance in advance of their meet and confers.[189]

185. Greer, *supra* note 1, at 6.

186. Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 192 (S.D.N.Y. 2012) (explaining "transparency in its proposed ESI search protocol made it easier for the Court to approve the use of [TAR]" because "such transparency allows the opposing counsel (and the Court) to be more comfortable with [TAR], reducing fears about the so-called 'black box' of the technology," and addressing concerns about "garbage in, garbage out" in training the tool.)

187. *Id.* (referring to fears about the so-called "black box" of TAR).

188. Greer, *supra* note 1, at 1.

189. *Id.* at 3, 6.

By addressing the disclosure in the ESI protocol in advance, parties can avoid the expense of litigating the dispute, and parties who truly do have concerns about producing their seed sets will not face uncertainty or possibly being forced to disclose.

Federal Rule of Civil Procedure 26(f) requires the parties to "confer as soon as practicable" about discovery, and particularly e-discovery issues.[190] Counsel should be both willing and prepared to discuss early on if and how they plan to proceed with the use of TAR. Agreement before proceeding is key because the cases to-date (addressing both the use of TAR generally as well as seed set disclosure in particular) only serve to highlight that the discovery disputes are even more contentious when parties have already invested time and resources and are faced with having to change horses midstream.[191] Counsel should be willing to address contentious issues because they will be no less contentious later on and these issues are "better to resolve up front"[192] than to litigate—enduring the cost and uncertainty—later. Equally important, for the meet and confer to be effective, parties need to be fully prepared which may require the involvement of e-discovery or IT consultants.[193] Judge Peck often asks parties to "bring your geek to court . . . ."[194]

Parties who remain uncomfortable with the disclosure of seed sets might consider negotiating other methods to assure the validity of the production and additionally should consider using the Continuous Active Learning methodology as it "eliminates issues about the seed set and stabilizing the TAR tool."[195] If there is next to no seed set, there is next to nothing to produce. However, while this may address the party's disclosure concern (which may be motivated either by a reticence to haggle on the particulars or work product concerns, or both) it does not address the arguably larger issue of ensuring the accuracy of a given TAR process and the substantive evidence it yields for the litigation at hand. Accordingly, where parties reach an agreement not to disclose their seed sets or where parties utilize methodologies with minimal seed sets, a magistrate judge may have a larger or earlier role to play (by looking at metrics of TAR defensibility other than seed sets) to ensure the methodology is sound and that the court is performing its *Daubert* gatekeeping function.

This Note recommends to courts that the parties' agreement as to whether seed sets will be disclosed should be a precursor to approving the use of TAR in the matter. This is the path the federal district court took in *Progressive Casualty Insurance Co. v. Delaney* finding that moving forward without an agreement

190. FED. R. CIV. P. 26(f).

191. Greer, *supra* note 1, at 6 ("Discovery disputes are most easily resolved at the beginning of a production and are most intractable after production has begun.").

192. Hon. Andrew J. Peck, *Introduction to eDiscovery*, *in* EDISCOVERY FOR CORPORATE COUNSEL § 1:6 (Carole Basri & Mary Mack eds., 2018) ("For counsel who feel they are educating the adversary or opening the door to contentious responding demands (you suggest 3 key player/custodians and 25 search terms, the adversary demands 10 custodians and 100 search terms), my response is: better to resolve that up front, by the Court if necessary.").

193. *Id.*

194. Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 193 (S.D.N.Y. 2012).

195. Hyles v. New York City, No. 10CIV3119ATAJP, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016).

"[would] only result in more disputes [and] further delay completion of discovery."[196]

While this Note does espouse the argument that judges have a *Daubert*-like gatekeeping function to play with respect to TAR, in the interest of judicial economy, the recommendations focus on addressing the "Garbage In, Garbage Out" problem proactively between the parties via disclosure of seed sets and ESI protocols. Some opponents of extending *Daubert* to the pre-trial use of TAR fear the time and expense associated with potential multi-day *Daubert* hearings,[197] however—aside from the fact that magistrate judges are already spending significant time presiding over these discovery disputes[198]—it is noteworthy that the recommendation of a *Daubert*-like analysis does not necessarily add a layer to the proceedings, but rather repositions one to where it is needed and will be most effective. If a party is using a faulty TAR methodology, it will necessarily affect the substantive evidence and its admissibility, and perhaps the development of the case, which will inevitably need to be addressed at some point. However, by positioning the gatekeeping-style analysis at the pre-trial stage when the TAR process is actually occurring, the court and parties are best situated to mitigate and minimize the effects of an unreliable method, and to course correct.

This Note envisions a magistrate's *Daubert*-like function with respect to TAR as *available* though not as the first cut at the issue. By recommending substantive haggling and agreement among the parties and their respective e-discovery professionals upfront, the magistrate judge is positioned not as a micromanager nor the primary line of defense, but rather more akin to a last line of defense (thus putting less strain on judicial resources).

If courts decide either that seed sets are non-relevant, or alternatively that even though seed sets are relevant they are nevertheless protected as attorney work product, the role of magistrate judges as gatekeepers will become even more necessary because as between themselves, parties will be left with few checks on the "Garbage In, Garbage Out" problem that is unique to TAR (as opposed to traditional manual review). If courts conclude that the Federal Rules do not empower them to require seed set disclosure, this Note recommends magistrate judges step in and play at least some baseline role in assessing the validity of a party's TAR process and the corresponding reliability and completeness of a production.

While traditional notions of relevancy may not require production of seed sets because the relevance/non-relevance coding is how the algorithm is, in essence trained, non-relevant documents have a relational relevance in a way that documents coded non-relevant per a traditional linear review do not. TAR is only as good as the human who trained the computer algorithm; it is

196. Progressive Cas. Ins. Co. v. Delaney, No. 2:11-CV-00678-LRH, 2014 WL 3563467, at *11 (D. Nev. July 18, 2014).

197. *See, e.g.*, Da Silva Moore v. Publicis Groupe, 2012 WL 1446534 at *2–3 (S.D.N.Y. Apr. 26, 2012) (questioning the potential applicability of the *Daubert* analysis).

198. John M. Facciola, Discovery: *Faster and Shorter*, 7 FED. CTS. L. REV. (2005), http://www.fclr.org/fclr/reviews/2005/fedctslrev7.shtml.

reasonable for parties to seek assurances that they are receiving reliable and complete productions in response to their discovery requests.[199] While the Federal Rules of Civil Procedure may not empower federal judges to compel the production of seed sets, at minimum judges can address the problem proactively by requiring agreement among the parties in the ESI protocol at the outset of the discovery process.

V. CONCLUSION

TAR is gaining wide acceptance in the e-discovery community and will likely become a necessary tool for litigators in the future, to the extent it is not already. However, the legal community needs a real and standardized avenue to check the potential "Garbage In, Garbage Out" problem of TAR—both to be confident in the resulting substantive evidence that serves the basis of a litigation, but also so parties are not forced to re-litigate the TAR issues anew at the outset of every matter.

The goal of litigation is "the just, speedy, and inexpensive determination of every action."[200] TAR is necessary to facilitate that goal.[201] Given the nature of the way attorneys and their clients communicate and conduct their business, the volume of electronically stored information is certain to continue to grow.[202] While a few million documents can be manually reviewed (though at great expense), eventually the number of documents responsive to discovery requests is likely to exceed the capabilities of manual review—at least on any reasonable timeline or budget.

Until courts begin to coalesce around hard and fast requirements, opposing attorneys should be prepared to discuss e-discovery protocols at the outset of the discovery to avoid litigating disputes. While at first glance the traditional applications of the Federal Rules of Civil Procedure and the Federal Rules of Evidence do not provide a clear hook for courts to require the disclosure of seed sets, the use of TAR in the document review process is itself a departure from tradition. And a necessary departure. Therefore, this Note argues seed sets have a relational or procedural relevance under the Federal Rules of Civil Procedure, though acknowledges Courts may nonetheless consider validation sets privileged as work product. Thus, this Note recommends parties address the disclosure of seed sets in their ESI protocols and Courts be prepared to engage in a *Daubert*-like analysis upon dispute to ensure parties are appropriately using TAR.

199. Goodman, *supra* note 12.

200. FED. R. CIV. P. 1.

201. *See supra* Section II.

202. *Guidelines for Addressing the Discovery of Electronically Stored Information*, US DIST. CT. OF. COLO., http://www.cod.uscourts.gov/Portals/0/Documents/Forms/CivilForms/E-Discovery_Guidelines.pdf (last visited Mar. 23, 2018).