Forthcoming in Volume 30, Issue 2 of the* Journal of Technology Law & Policy*, University of Florida Levin College of Law. This excerpt is subject to revision and editorial changes prior to final publication.

ARTIFICIAL INTELLIGENCE IN DOCUMENT REVIEW: STATISTICALLY DEFENDING THE PROCESS AND RESULTS OF E-DISCOVERY WORKFLOWS

Aron Ahmadia, *Nathan Reff*, *Matthew Reiber*, *Chad Roberts*, & *Cristin Traylor*

Abstract

The use of artificial intelligence for document review allows legal teams to process and produce electronic documents at a scale, speed, and cost that would have been unheard of only a few years ago.

But as legal teams become increasingly dependent on artificial intelligence to locate and prioritize responsive documents, courts and their litigation opponents rightly demand heightened verification that the workflow is sufficiently well designed, and the evidentiary output sufficiently accurate, to ensure a full and fair adjudication of the facts at trial. This Article explores the processes and statistical tools available to litigants to respond to these sorts of objections and concerns, focusing on the methods available to measure the completeness and accuracy of an e-discovery workflow. Drawing on principles of information retrieval and machine learning, the Article examines common metrics and explains their application within the framework of a defensible e-discovery workflow. This Article also addresses statistical sampling techniques, including random and stratified sampling, and discusses how confidence intervals and margins of error can be used to assess the reliability of the AI-driven output. This Article gives particular attention to the validation of technology-assisted review protocols, including continuous learning, simple learning, and other iterative workflows that adapt to human input and subsequent machine analysis. By examining both quantitative approaches and qualitative safeguards, and drawing on interdisciplinary literature from academia, industry, and government, this Article evaluates each technique's strengths, limitations, and situational suitability—thus providing courts and practitioners with an actionable guide for selecting defensible validation strategies across diverse review contexts. This Article provides relevant stakeholders with a better understanding of how to evaluate and measure the performance of an AI-driven workflow and verify whether it meets the requirements of the Federal Rules of Civil Procedure. This Article is particularly helpful because it provides a roadmap for litigants and judges who must confront AI-related issues at critical junctures in the litigation stream. And this Article is particularly

timely because continued advances in artificial intelligence (particularly generative artificial intelligence) will bring this tool to a broader range of litigants who will use it and judges who must evaluate its merits.

INTRODUCTION

One of the more significant challenges confronting civil litigants is marshaling ever-increasing volumes[1] of electronic evidence[2] through the discovery process.[3] While the sources of this evidence are often readily discernible, reviewing it prior to production can be burdensome and costly.[4] This is not a new problem, of course,[5] and the Federal Rules of Civil Procedure have been amended several times since the early 1980s

---

1. *See, e.g.*, *In re* Aspartame Antitrust Litig., 817 F. Supp. 2d 608, 614 (E.D. Pa. 2011) (defendants collectively gathered the equivalent of eighty million pages of potentially responsive documents); Tampa Bay Water v. HDR Eng'g, Inc., No. 08-CV-2446, 2012 WL 5387830, at * 21 (M.D. Fla. Nov. 2, 2012) ("This was a lengthy, highly technical case which involved 17 million pages of documents."); *In re* Online DVD-Rental Antitrust Litig., 779 F.3d 914, 925 (9th Cir. 2015) ("Netflix ultimately produced almost 15 million pages . . . ."); *In re* Allergan Biocell Textured Breast Implant Prods. Liab. Litig., MDL No. 2921, 2022 WL 16630821, at *1 (D.N.J. Oct. 25, 2022) ("Defendants . . . collected and indexed 9.371 terabytes of data . . . ."); Fast Memory Erase, LLC v. Spansion, Inc., No. 10-cv-0481, 2010 WL 5093945, at *4 (N.D. Tex. Nov. 10, 2010) (Intel defendants collected and processed "more than 2,100 gigabytes of electronically stored information."); Brief of Defendants-Appellees Cisco Ironport Systems, LLC and Return Path, Inc., CBT Flint Partners, LLC v. Return Path, Inc., 737 F.3d 1320 (Fed. Cir. 2013) (No. 2013-1036), 2013 WL 1741921, at *2 (Cisco collected "1.2 terabytes of data – or the equivalent of 12% of the printed volume of the Library of Congress.").

2. The increasing volume is driven largely by the proliferation of unstructured data. Unstructured data is free-form data "that either does not have a data structure or has a data structure not easily readable by a computer without the use of a specific program designed to interpret the data; created without limitations on formatting or content by the program with which it is being created." *The Sedona Conference Glossary: eDiscovery & Digital Information Management, Sixth Edition*, 27 SEDONA CONF. J. 1, 131–32 (2026) [hereinafter *The Sedona Conference Glossary*]. Common examples of unstructured data include email, text messages, social media posts, audio files, and videos. *See, e.g.*, KLDiscovery, *Unstructured Data: The Black Hole of Ediscovery*, https://www.kldiscovery.com/blog/unstructured-data-the-black-hole-of-ediscovery [https://perma.cc/TQN9-8QVY].

3. The obligation to marshal data through the discovery process is most frequently encountered in the context of a response to a request for the production of documents and electronically stored information pursuant to Rule 34 of the Federal Rules of Civil Procedure. FED. R. CIV. P. 34(a)(1)(A). The term "electronically stored information" refers to "information that is stored electronically, regardless of the media or whether it is in the original format in which it was created . . . ." *The Sedona Conference Glossary*, *supra* note 2, at 45–46.

4. *See, e.g.*, Lawson v. Spirit AeroSystems, Inc., No. 18-1100, 2020 WL 6343292, at *1, 1–2 (D. Kan. Oct. 29, 2020) (collection of 322,524 electronic documents resulted in the production of approximately 24,000 at an end-to-end cost of $791,700.21).

5. The antitrust battles of the 1970s were the first to confront large-scale document discovery. *See, e.g.*, IBM Corp. v. United States, 471 F.2d 507, 523 (2d Cir. 1972) (Mulligan, J., dissenting) ("It is indeed mind-boggling to contemplate [the production of] 17 million document pages which in bulk weigh 87 tons and would stretch from coast to coast.").

to make discovery more efficient and less expensive.[6] But even the best procedural system—with the most advanced rules—will be tested from time to time as individuals and entities create more electronic information, which is stored in a wider array of locations, and that is discoverable in a broader range of cases.[7] Rule 26 laudably defines the *scope* of discovery in relation to the needs of the case (as well as relevance and privilege) and thus allows the breadth of discovery to ebb and flow from case to case.[8] But Rule 26 does not address the *success* of

---

6. Since the late 1970s, the scope of discovery has been a recurring issue for the Advisory Committee on Civil Rules, the subject matter committee of the Judicial Conference tasked with responsibility for studying and recommending changes to the Federal Rules of Civil Procedure. The Advisory Committee's work resulted in several amendments to Rule 26 and the scope of discovery. *See* FED. R. CIV. P. 26(b)(1)(iii) (1983) (authorizing court to limit the frequency and extent of discovery when it is "unduly burdensome or expensive, taking into account the needs of the case, the amount in controversy, limitations on the parties' resources, and the importance of the issues at stake in the litigation"); FED. R. CIV. P. 26(b)(1)–(2) (1993) (rearranging content of Rule 26(b) by moving specific limitations on discovery from (b)(1) to (b)(2) and authorizing court to limit discovery when "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 discovery in resolving the issues"); FED. R. CIV. P. 26(b)(1) (replacing "relevant to the subject matter involved in the pending action" with "the claim or defense of any party," although authorizing district court to allow broader subject matter discovery upon a showing of good cause); FED. R. CIV. P. 26(b)(2)(B) (2006) (adding additional limitation on scope of discovery with respect to electronically stored information that is not reasonably accessible because of undue burden or cost absent a court order); FED. R. CIV. P. 26(b)(1)–(2) (2015) (rearranging the text regarding the scope of discovery, and the limitations on discovery, by moving certain text from Rule 26(b)(2) to Rule 26(b)(1) and providing that "the scope of discovery" is defined in relation to relevance and privilege, as well as proportionality 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 the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit").

7. George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?*, 13 RICH. J. L. TECH. 10, 13 (2007) ("[T]he amount of information in business has increased by thousands, if not tens of thousands of times in the last few years. In a small business, whereas formerly there was usually one four-drawer file cabinet full of paper records, now there is the equivalent of two thousand four-drawer file cabinets full of such records, all contained in a cubic foot or so in the form of electronically stored information. This is a sea change.").

8. Rule 26(b)(1) provides that parties may obtain discovery "regarding any nonprivileged matter that is relevant to any party's claim or defense and proportional to the needs of the case . . . ." FED. R. CIV. P. 26(b)(1). The extent to which discovery is "proportional to the needs of the case" is measured against six specific factors: "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 the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit." FED. R. CIV. P. 26(b)(1). Several state procedural systems similarly define the scope of discovery with respect to both relevance and proportionality. *See, e.g.*, AZ. R. CIV. P. 26(b)(1) (scope of discovery is identical to that in Federal Rule 26); DEL. CH. CT. R. 26(b)(1) (scope of discovery is substantively the same as defined in Federal Rule 26).

discovery and, specifically, whether it unearthed an optimal corpus of evidence from which the parties can support their factual contentions at trial—optimal in the sense of completeness and accuracy after considering the cost and benefit of doing more or better.[9] To compound this situation, the federal courts have not addressed what constitutes successful e-discovery with sufficient frequency, or at a meaningful level of generality,[10] to provide a measuring stick against which litigants can evaluate their own efforts and courts can adjudicate claims of inadequacy or insufficiency. The scholarship regarding successful discovery is equally sparse.[11]

An evolving collection of artificial intelligence technologies, however, offers the possibility of achieving successful e-discovery by maximizing completeness and accuracy while minimizing burden and cost, even when applied to the largest data collections.[12] These

9. Completeness or "recall" refers to the percentage of responsive items within a document population that are correctly identified as such and thereafter produced in discovery. Accuracy or "precision" in turn refers to the percentage of items retrieved by a particular search methodology that are in fact responsive. *See, e.g.*, Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 189 (S.D.N.Y. 2012) ("Recall is the fraction of relevant documents identified during a review; precision is the fraction of identified documents that are relevant."); Alivecor, Inc. v. Apple, Inc., No. 21-cv-03958, 2023 WL 2224431, at *1 (N.D. Cal. Feb. 23, 2023) (same).

10. *See, e.g.*, *In re*: Insulin Pricing Litig., No. 23-md-3080, 2025 WL 1112837, at *5–7 (D.N.J. Apr. 11, 2025) (describing validation protocol); *In re*: Uber Techs., Inc., Passenger Sexual Assault Litig., MDL No. 3084, 2024 WL 3491760, at *5–6 (N.D. Cal. Mar. 15, 2024) (requiring disclosure of recall and prevalence but declining to specify a percentage for either); *In re*: Social Media Adolescent Addiction/Personal Injury Prods. Liab. Litig., No. 22-md-03047, 2024 WL 1786293, at *7 (N.D. Cal. Feb. 20, 2024) (ordering that each party "shall take reasonable steps to validate its review process" but not specifying them); BCBSM, Inc. v. Walgreen Co., No. 20-cv-01853, 2023 WL 6852533, at *8–10 (N.D. Ill. July 21, 2023) (providing validation worksheet "to obtain an estimate of the quantity and quality (*i.e.*, materiality) of responsive documents identified and missed by a particular review effort, and to suggest the source of any inadequacy that may be indicated by such estimate"); Deal Genius, LLC v. O2COOL, 682 F. Supp. 3d 727, 734-35 (N.D. Ill. July 14, 2023) (describing use of elusion testing as a validation metric); *In re* Diisocyanates Antitrust Litig., Misc. No. 18-1001, 2021 WL 4295729 (W.D. Pa. Aug. 23, 2021) (evaluating parties' competing approaches to validation); City of Rockford v. Mallinckrodt ARD Inc., 326 F.R.D. 489, 492–95 (N.D. Ill. 2018) (describing utility of elusion test to validate adequacy of production); *In re* Broiler Chicken Antitrust Litig., No. 16-cv-08637, at *4–6 (N.D. Ill. Jan. 3, 2018) (identifying validation protocol "to ensure a reasonable production"); Rio Tinto PLC v. Vale, No. 14 Civ. 3042, 2015 WL 13956130, at *12 (S.D.N.Y. Sep. 8, 2015) (establishing validation protocol to determine whether the parties' respective productions were adequate).

11. *See, e.g.*, Karl Schieneman and Thomas C. Glick III, *The Implications of Rule 26(g) on the Use of Technology-Assisted Review*, 7 FED. CTS. L. REV. 239, 269–73 (2013) (discussing the use of recall and validation statistics to satisfy the requirements of Federal Rule 26(g)); Marua R. Grossman & Gordon V. Cormack, *Comments on "The Implications of Rule 26(g) on the Use of Technology Assisted Review*," 7 FED. CTS. L. REV. 285, 301–13 (2014) (questioning the utility of relying on recall to evaluate the success of discovery).

12. *See* Maura R. Grossman, Gordon V. Cormack, & Jason R. Baron, *Does the LLMperor Have New Clothes? Some Thoughts on the Use of LLMs in Ediscovery*, 109 THE ADVOC. (TEXAS) 22, 23 (2024) ("Over the past 15 years or so, the legal profession has become increasingly aware

technologies—which predict the responsiveness of individual documents—allow human reviewers to focus on the responsive documents in a collection and thus avoid the often-greater number of nonresponsive ones. The resulting reduction in human involvement lowers the cost of discovery,[13] accelerates its completion, and shortens the time to trial, thus promoting the "just, speedy, and inexpensive determination of every action and proceeding."[14]

Yet while litigants have used artificial intelligence in discovery for more than a decade,[15] disputes regarding its ability to reliably

---

of the availability of various forms of AI used specifically to find responsive documents in complex litigation. The two most-established methods – commonly dubbed 'TAR 1.0' and 'TAR 2.0' – employ supervised machine learning to distinguish responsive documents from non-responsive documents."). *See also* Lea Malani Bays, *Harness the Power of Generative AI in Document Review*, 61- TRIAL MAG. 34, 37 (May 2025) ("GenAI is the new hope for solving e-discovery challenges. . . . [I]nitial testing shows promising results, with generative AI identifying relevant documents with increased accuracy. Early studies suggest that GenAI can identify over 95% of relevant documents in a dataset and may even outperform both human reviewers and TAR.").

13. Publicly available information regarding the burden and expense of attorney review of electronic evidence prior to production is limited. *Lawson v. Spirit AeroSystems, Inc.*, No. 18-1100, 2020 WL 6343292 (D. Kan. Oct. 29, 2020) is one of the few cases to address the issue. In that case, the district court awarded $398,001.26 for first and second level attorney review based on 3,620.9 hours. Publicly available information regarding the cost of storing or "hosting" electronic evidence during review is more readily available. *See, e.g.*, United States *ex rel.* King v. Solvay S.A., Civil Action H-06-2662, 2016 WL 3523873, at *15 (S.D. Tex. June 28, 2016) ($429,964.95 for data storage); Abbott Point of Care, Inc. v. Epocal, Inc., No. CV-08-S-543, 2012 WL 7810970, at *2 (N.D. Ala. Nov. 5, 2012) ($340,498 total for hosting); Murphy v. Precision Castparts Corp., No. 3:16-cv-00521, 2021 WL 4524153, at *6 (D. Or. Oct. 4, 2021) ($168,069 for data hosting).

14. FED. R. CIV. P. 1. The reduction of human involvement should also reduce the frequency of disputes regarding the proportionality of particular discovery by eliminating a common justification for reducing the breadth of discovery in particular cases. *See, e.g.*, Ruggles v. WellPoint, Inc., No. 08-CV-201, 2010 WL 11570681, at *5 (N.D.N.Y. Dec. 28, 2010) ("Ever since the advent of storing monumental quantities of information and data electronically and its dynamic impact upon discovery, by all measures, the cost of producing e-discovery, which is generally borne by the responding party, consumes an inordinate quantum of most discovery debates. Here, the cost of e-discovery dominates our discussion.").

15. *See, e.g.*, Roshanak Omrani et al., *Beyond the Bar: Generative AI as a Transformative Component in Legal Document Review*, RELATIVITY ODA LLC (2024) at 2 https://resources.relativity.com/rs/447-YBT-249/images/Beyond-the-Bar_preprint_2-9%20%281%29.pdf?version=0 [https://perma.cc/Z5R2-BWSN] ("Supervised machine learning first came into use in eDiscovery around 2005, with courts in the United States, England, Ireland, Australia and other jurisdictions explicitly encouraging its use starting in 2012."). *See also In re* Valsartan, Losartan, & Irbesartan Prods. Liab. Litig., 337 F.R.D. 610 (D.N.J. 2020) ("We are past the time when parties and courts view TAR as an outlier."). For early cases involving the use of technology in the discovery process, see Da Silva Moore v. Publicis Groupe, 287 F.R.D. 182, 183 (S.D.N.Y. 2012) ("This judicial opinion now recognizes that computer-assisted review is an acceptable way to search for relevant ESI in appropriate cases."); Glob. Aerospace Inc. v. Landow Aviation, L.P., No. CL 61040, 2012 Va. Cir. LEXIS 50, at *2 (Va. Cir. Ct. Apr. 23, 2012) ("[I]it

differentiate between responsive and nonresponsive documents continue to occur and recur,[16] regardless of the expertise of the lawyers in the case.[17] The requesting party contends that the process was not adequately executed, that the resulting production is incomplete, and that the requirements of Rules 26(b)(1)[18] and 26(g)(1)[19] are therefore not

---

is hereby ordered Defendants shall be allowed to proceed with the use of predictive coding for purposes of the processing and production of electronically stored information . . . ."); W. Penn Allegheny Health Sys., Inc., No. 09-cv-0480, 2013 WL 12134, at *1 (W.D. Pa. Feb. 28, 2013) (authorizing use of "a technology-assisted review process, including, for example, one that predictively codes documents as subject to or not subject to production . . . ."); W Holding Co. v. Chartis Ins. Co. of Puerto Rico, No. 11-2271, 2013 WL 1352562, at *5 (D.P.R. Apr. 3, 2013) ("The Parties shall meet and confer in good faith about any other technology or process that a producing party proposes to use to streamline the culling, review and production of ESI (e.g., email threading, near de-duplication, technology assisted review."); Edwards v. Nat'l Milk Producers Fed'n, No. 11-CV-04766, at *1 (N.D. Cal. Apr. 17, 2013) (authorizing use of predictive coding "for a more cost efficient and higher quality review"); United States v. Educ. Mgmt. LLC, No. 07-cv-00461, 2013 WL 12140442, at *8 (W. D. Pa. Nov. 24, 2014) (approving use of computer assisted review); Rio Tinto PLC v. Vale S.A., 306 F.R.D. 125, 127 (S.D.N.Y. 2015) ("[I]t is now black letter law that where the producing party wants to utilize [technology assisted review] for document review, courts will permit it.").

16. *See, e.g.*, *In re*: Insulin Pricing Litig., No. 23-md-3080, 2025 WL 1112837, at *5 (D.N.J. Apr. 11, 2025) ("The parties do not dispute that there must be a validation process for the TAR model. However, they dispute the scope of the data set that would comprise the validation process."); *In re*: Uber Techs., Inc., Passenger Sexual Assault Litig., MDL No. 3084, 2024 WL 3491760, at *5 (N.D. Cal. Mar. 15, 2024) (parties agreed defendant would disclose recall and prevalence but disagreed about whether defendant would also have to disclose "the input quantities used to calculate" them); *In re*: Social Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig., No. 22-md-03047, 2024 WL 1786293, at *7 (N.D. Cal. Feb. 20, 2024) ("In general, the Parties agree that some form of validation of the review procedures is appropriate. The disputes center on how much information to share about validation and the burden required to seek additional information from a Producing Party."); Deal Genius, LLC v. O2Cool, LLC, 682 F. Supp. 3d 727, 738–39 (N.D. Ill. 2023) (reliability of elusion testing called into question based on the results from the application of an additional search term against the null set); *In re* Diisocyanates Antitrust Litig., Misc. No. 18-1001, 2021 WL 4295729, at *2–5 (W.D. Pa. Aug. 23, 2021) (dispute regarding particular methodology for performing elusion test); City of Rockford v. Mallinckrodt ARD Inc., 326 F.R.D. 489, 491–92 (N.D. Ill. 2018) (resolving dispute regarding whether defendant would be required to conduct elusion test at conclusion of its document review or whether plaintiff would be required to identify potential gaps and omissions in document production).

17. *See, e.g.*, *In re:* Insulin Pricing Litig., No. 23-md-3080, (D.N.J. 2023) (Davis Polk; Jones Day; Kirkland & Ellis); *In re* Diisocyanates Antitrust Litig., Misc. No. 1001 (W.D. Pa. 2018) (Eckert Seamans, Mayer Brown; Morgan Lewis; Reed Smith); *In re* Broiler Chicken Antitrust Litig., No. 16-cv-08637 (N.D. Ill. 2016) (Simpson Thacher; Skadden Arps; White & Case).

18. *See supra* note 8.

19. Rule 26(g)(1) imposes a signature requirement on all discovery responses, and that signature carries with it deemed certifications that the response is "consistent with these rules" and is not "unreasonable . . . considering the needs of the case, prior discovery in the case, the amount in controversy, and the importance of the issues at stake in the litigation." FED. R. CIV. P. 26(g)(1). These deemed certifications are best understood in relation to the evolving norms of the

satisfied. The producing party counters that the process was sound, that the resulting production is proportionate to the needs of the case, and that, in any event, it is best situated to evaluate the "procedures, methodologies, and technologies appropriate for preserving and producing" its own information.[20] The court must then sift through arguments from lawyers and testimony from experts to craft an appropriate resolution.

This Article attempts to reduce the frequency of these disputes, as well as facilitate their resolution when they cannot be avoided, by providing litigants and judicial officers with a thorough, readily accessible explanation of the use of artificial intelligence in discovery and the statistical measures that can be used to validate its successful performance (or alternatively, identify its shortcomings) regardless of the selected workflow. Part I describes the evolution of artificial intelligence in e-discovery. This part highlights how technology can be deployed to reduce the burden and expense of e-discovery while improving the quality of the output. Part II then provides a "user guide" of sorts for validating e-discovery workflows and a step-by-step process for selecting and implementing the most appropriate methodology for individual cases. While this part acknowledges that perfection is not attainable,[21] it demonstrates that a well-designed workflow, coupled with the use of appropriate technology, increases the likelihood that e-discovery will be successful and that the "truth will out."

---

profession regarding discovery obligations. *See, e.g.*, Karl Schieneman & Thomas C. Gricks III, *The Implications of Rule 26(g) on the Use of Technology Assisted Review*, 7 FED. CTS. L. REV. 239, 246–47 (2013) ("[T]he producing party must conduct a reasonable, good faith, and diligent search for responsive ESI, recognizing that the objective is not perfection. That obligation, however, is tempered by the proportionality concept . . . and the ultimate goal of 'secur[ing] the just, speedy and inexpensive determination of every action and proceeding' . . . . ") (quoting FED. R. CIV. P. 1).

20. *In re* Insulin Pricing Litig., No. 23-md-3080, 2025 WL 1112837, at *1 (D.N.J. Apr. 11, 2025) (quoting *The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production*, 19 SEDONA CONF. J. 1, 52, 118 (2018)). *See also* Nichols v. Noom Inc., No. 20-CV-3677, 2021 WL 948646, at *2 (S.D.N.Y. Mar. 11, 2021) ("[A] producing party is best situated to determine its own search and collection methods so long as they are reasonable.").

21. *See, e.g.*, Fed. Hous. Fin. Agency v. HSBC N. Am. Holdings Inc., No. 11 Civ. 6189, 2014 WL 584300, at *2 (S.D.N.Y. Feb. 14, 2014) ("Parties in litigation are required to be diligent and to act in good faith in producing documents in discovery. The production of documents in litigation such as this is a herculean undertaking . . . . [N]o one could or should expect perfection from this process. All that can be legitimately expected is a good faith, diligent commitment to produce all responsive documents uncovered when following the protocols to which the parties have agreed, or which a court has ordered.").