



Artificial Intelligence in Defense Practice Seminar

March 10, 2026

PAPER TITLE: E-Discovery

Brett Creasy

bit-x-bit LLC

501 Grant Street, Suite 875

Pittsburgh, PA 15219

brett.creasy@bit-x-bit.com

Chris Jacobs

Houston Harbaugh PC

401 Liberty Avenue, Floor 22

Pittsburgh, PA 15222-1005

jacobscm@hh-law.com

Acacia Perko

Houston Harbaugh PC

401 Liberty Avenue, Floor 22

Pittsburgh, PA 15222-1005

perkoab@hh-law.com

John Unice

bit-x-bit LLC

501 Grant Street, Suite 875

Pittsburgh, PA 15219

john.unice@bit-x-bit.com

SESSION TITLE: *Discovery in the AI Era: Forensic Data Retrieval; Advanced E-Discovery Practices, from Active Learning to Generative AI*

Presented by:

Brett Creasy, *bit-x-bit LLC, Pittsburgh, PA*

Chris Jacobs, *Houston Harbaugh PC, Pittsburgh, PA*

Acacia Perko, *Houston Harbaugh PC, Pittsburgh, PA*

John Unice, *bit-x-bit LLC, Pittsburgh, PA*

Brett Creasy directs the company's overall operation in digital forensics, eDiscovery, cybersecurity, and incident response at bit-x-bit, LLC. Providing exceptional expertise and guidance to clients, Brett draws on more than two decades of experience in cybersecurity and digital forensics. He regularly serves as an expert on matters involving the theft of trade secrets

and confidential information, fraud, recovery of cell phone text messages, and the alteration of electronic evidence and documents.

Brett has conducted hundreds of forensic examinations for cases involving intellectual property theft, unauthorized use of computer systems, harassment, computer hacking, criminal and white-collar crimes, and employment class action litigation, as well as arbitration proceedings and internal investigations. He has been designated as an expert in state and federal courts throughout the country and has testified in numerous cases, most notably those involving the theft of confidential and trade secret information.

Brett's extensive experience also includes handling a wide range of cybersecurity and incident response matters, including hacking attacks involving malware, and personally identifiable information (PII) and protected health information (PHI). His ability to manage fast-paced engagements has made him a trusted resource for organizations facing critical digital threats.

Christopher M. Jacobs is a litigation attorney focusing on insurance coverage and insurance bad faith matters and business litigation. He represents national insurers in complex bad faith and coverage matters and his business litigation practice includes U.S.-based and international commercial disputes, trade secret and DTSA claims, and trademark infringement. Christopher also advises clients on emerging legal issues involving artificial intelligence and algorithmic decision-making and is an active DRI committee member and speaker.

Acacia B. Perko is a trial lawyer focusing on business litigation, intellectual property, employment disputes, and insurance coverage and defense matters, including trade secrets and AI-related issues. She represents clients in state and federal courts. Acacia serves as DRI's Trade Secrets Substantive Law Group Chair for the Intellectual Property Litigation Committee and has been recognized as a Super Lawyers "Rising Star" since 2017.

John Unice directs the business, strategic, and overall operations of bit-x-bit. Providing expectational high-level expertise and guidance to bit-x-bit's clients, John draws on two decades of experience as chief corporate counsel and as a trial lawyer in major corporate and large law firm settings.

John's wide-ranging portfolio includes litigating sophisticated cases, utilizing complex electronic evidence and leading digital risk-mitigation initiatives, both in the U.S. and internationally. Before joining bit-x-bit, John was Assistant Secretary and a senior legal advisor at the Pittsburgh headquarters of Covestro LLC and, prior, at Bayer Corporation, advancing his clients' interests in highly regulated industries. At Covestro, he oversaw legal affairs for the company's U.S.-based polycarbonate business and held global responsibility for the company's litigation docket. John also led and collaborated on a range of strategic e-discovery, cybersecurity, and digital records management initiatives to mitigate corporate risk and win cases.

As associate with the Trial Practice Group at Jones Day (Pittsburgh), John's savvy use of e-discovery and digital forensics helped him secure favorable outcomes in numerous multi-million-dollar cases for both the defense and the prosecution.

Leveraging Artificial Intelligence to Streamline Factual Investigation and Case Development

A Practical and Defensible Framework for Modern Litigation

I. Introduction: Why AI in Litigation Is No Longer a Novelty

Artificial intelligence is no longer a speculative or emerging concept in litigation. Courts across jurisdictions have now spent more than a decade addressing technology-assisted review, analytics-driven discovery workflows, and the proportional use of advanced tools to manage ever-expanding volumes of electronically stored information (“ESI”). What has changed in recent years is not whether technology may be used, but how it is deployed, explained, and defended.

Modern litigation increasingly presents a familiar problem: massive data sets, compressed timelines, escalating costs, and heightened judicial expectations regarding efficiency and cooperation. Against that backdrop, AI-enabled tools—ranging from traditional analytics to large language models—are being incorporated into litigation workflows. The central question for practitioners is no longer whether these tools are permissible, but whether their use is reasonable, transparent, and defensible under the Federal Rules of Civil Procedure and analogous state rules.

This shift mirrors broader changes in judicial management of civil litigation. Judges increasingly expect parties to articulate *how* discovery decisions were made, *why* particular methodologies were selected, and *what steps were taken to ensure reliability*. As discovery disputes have moved away from debates over theoretical completeness and toward proportionality-driven outcomes, courts have shown a growing willingness to approve innovative workflows—provided counsel can explain and defend them.

Importantly, AI adoption does not occur in a doctrinal vacuum. The Federal Rules of Civil Procedure are technology-neutral by design. Rule 1’s mandate of “the just, speedy, and inexpensive determination of every action” and Rule 26’s proportionality framework implicitly encourage the use of tools that advance those goals. AI therefore fits naturally within existing procedural doctrine rather than requiring new rules or judicial reinvention.

This manuscript examines AI in litigation not as a disruptive force, but as part of a continuum of discovery practices already accepted by courts. When properly implemented, AI tools shift attorneys away from time-intensive, low-value document review tasks and toward higher-order legal analysis, strategic decision-making, and early case assessment—without undermining professional judgment or ethical obligations.

II. From Analytics to GenAI: A Continuum Rather Than a Leap

Courts have long recognized that manual, linear document review is neither the gold standard nor a required one. Early judicial acceptance of analytics and technology-assisted review (“TAR”) rested on a practical reality: humans are inconsistent, expensive, and error-prone when reviewing large document sets, particularly under time pressure.

Discovery analytics generally fall into two broad categories. **Conceptual analytics** include clustering, concept searching, and visualization tools designed to identify thematic relationships among documents. **Structured analytics** include email threading, near-duplicate detection, language identification, and automated redactions. These tools do not decide legal issues; they assist counsel in organizing and prioritizing information.

Courts evaluating early analytics tools quickly recognized that these technologies did not replace legal judgment; instead, they improved consistency and efficiency at scale. Traditional keyword searching, while familiar, often suffered from over-inclusiveness, under-inclusiveness, and inconsistent application across reviewers. Analytics tools addressed these weaknesses by identifying patterns that humans frequently miss, particularly in large, unstructured datasets.

TAR workflows evolved from early “seed set” models (often referred to as TAR 1.0) to continuous active learning (“CAL” or TAR 2.0), in which models are trained dynamically as reviewers code documents. Courts evaluating these technologies focused on recall, precision, validation, and proportionality—not on technical novelty.¹

The evolution from static seed-set training to continuous active learning marked an important doctrinal moment. CAL systems allowed review strategies to adapt in real time as new information emerged, aligning discovery practice more closely with Rule 26’s proportionality mandate. Rather than front-loading costs to achieve theoretical completeness, parties could assess marginal utility as review progressed and make informed decisions about when continued review no longer justified its expense.

More recent generative AI tools, including large language models, should be understood as an extension of these principles rather than a departure from them. While large language models introduce new capabilities—such as semantic summarization, chronology building, and multi-issue reasoning—they do not alter the fundamental legal inquiry. Courts remain concerned with whether discovery efforts are reasonable in scope, transparent in execution, and proportionate to the needs of the case. GenAI does not displace these standards; it operates within them.

¹ *Recall* measures the **thoroughness** of the search. It asks: Of all the relevant documents that exist in the entire collection, what percentage did the AI actually find? *Precision* measures the **efficiency** of the search. It asks: Of all the documents the AI flagged as “relevant,” how many are actually relevant? *Validation* is the **quality control** process used to prove that the TAR model actually worked. *Proportionality* is what courts consider under Fed R. Civ. P. 26(b)(1) and state equivalents that balances the need for information against the cost of getting it.

III. Judicial Acceptance: What Courts Actually Care About

A. Reasonableness as the Governing Standard

Judicial acceptance of advanced analytics is well established. In *Da Silva Moore v. Publicis Groupe*, one of the earliest decisions addressing predictive coding, the court emphasized that computerized review methods can be “at least as accurate, if not more so, than manual review,” and that the Federal Rules “do not require perfection.”² This recognition has become foundational to modern discovery jurisprudence.

Across jurisdictions, courts have repeatedly emphasized that discovery obligations are governed by reasonableness rather than perfection. No review methodology—human or machine—can guarantee error-free results, particularly when millions of documents are involved. Courts therefore evaluate whether a party conducted a reasonable inquiry, not whether it captured every potentially responsive document.

Subsequent decisions have reinforced several consistent themes. Producing parties are generally entitled to choose their discovery methodology, provided it is reasonable. Validation and quality control matter, particularly where large volumes of data are involved. Transparency facilitates cooperation, but does not require disclosure of proprietary algorithms or internal tooling. Discovery disputes are evaluated holistically, focusing on process rather than outcomes.

For example, in *Global Aerospace Inc. v. Landow Aviation, LP*, the Virginia Circuit Court observed that analytics “is capable of locating upwards of seventy-five percent of the potentially relevant documents . . . at a fraction of the cost and in a fraction of the time of linear review.”³

And, in *Federal Housing Finance Agency*, the Southern District of New York permitted the Defendants to use predictive coding over plaintiffs objections.⁴ The Court observed: “the literature that the Court reviewed . . . indicate[s] that predictive coding has a better track record in the production of responsive documents than human review.” Further, the Court explained “[N]o one could or should expect perfection from the discovery process. All that can be legitimately expected is a good faith . . . commitment to produce . . . responsive documents.”

The Western District of Pennsylvania provided a detailed decision describing how courts evaluate TAR workflows. The Special Master in the *In re Diisocyanates Antitrust Litigation* emphasized that “transparency transcends cooperation” and requires disclosure sufficient to make meet-and-confer discussions meaningful.⁵ At the same time, the court rejected rigid or formulaic stopping rules and reaffirmed that the legal standard remains reasonableness—not exhaustive completeness.

² *DaSilva Moore v. Publicis Groupe*, 1:2011cv01279 (S.D.N.Y. April 25, 2012).

³ *Global Aerospace Inc. v. Landow Aviation, L.P.*, 2012 WL1431215, No. CL 61040 (Va. Cir. Ct. Apr. 23, 2012).

⁴ *Federal Housing Finance Agency v. HSBC North America Holdings, Inc.*, 2014 WL584300 (S.D.N.Y. Feb. 14, 2014)

⁵ *In re Diisocyanates Antitrust Litig.*, No. MC 18-1001, 2021 WL 4295729, at *7 (W.D. Pa. Aug. 23, 2021), report and recommendation adopted, No. MC 18-1001, 2021 WL 4295719 (W.D. Pa. Sept. 21, 2021).

Significantly, the *Diisocyanates* court examined not only the existence of a TAR protocol, but also how that protocol accounted for validation, recall estimation, and iterative testing. The court’s analysis reflects an understanding that discovery methodologies must be evaluated in context, with attention to the nature of the case, the volume of data, and the significance of remaining unreviewed material.

For example, a review process that continues to surface duplicative invoices or low-level administrative communications may reasonably conclude earlier than one that continues to identify novel or highly probative documents. Courts have consistently recognized that such qualitative assessments are a legitimate component of proportionality analysis.

Similarly, in *Huntsman v. Southwest Airlines Co.*, the court approved a hybrid approach using keyword searches in tandem with technology-assisted review, holding that such a workflow satisfied the obligation to conduct a reasonable inquiry under Rule 26.⁶ The decision underscores that proportionality, phased discovery, and defensible methodology are often more important than any particular technical choice.

IV. Agentic AI as Workflow Orchestration, Not Autonomous Judgment

Recent discussion of “agentic AI” has generated concern that AI systems are making independent decisions or replacing attorney judgment. In litigation practice, however, agentic AI is best understood as workflow orchestration rather than autonomy.

Much of the concern surrounding agentic AI arises from terminology rather than functionality. In litigation contexts, “agents” do not act independently or exercise discretion. Instead, they perform narrowly defined tasks within parameters established, and quality-checked, by counsel. Each task—whether identifying documents tied to a witness, assembling chronologies, or flagging inconsistencies—is initiated, reviewed, and validated by human attorneys.

Unlike traditional TAR workflows, agentic AI tools do not require seed sets or wholesale retraining when new issues emerge. Counsel may instead investigate new questions incrementally and surgically, reducing cost while preserving proportionality. Importantly, this does not eliminate the need for methodology; it changes its form.

This incremental approach aligns closely with modern discovery principles. Rather than committing to expansive, front-loaded review protocols, counsel may adjust scope based on newly learned facts. From a proportionality standpoint, this allows discovery to remain responsive to the actual needs of the case rather than hypothetical worst-case scenarios.

⁶ “Southwest’s approach to using keyword searches and technology-assisted review in tandem does not offend the court’s expectation that the parties conduct a reasonable inquiry as required by the rules.” *Huntsman v. Sw. Airlines Co.*, No. 19-CV-00083-PJH, 2021 WL 3504154, at *3 (N.D. Cal. Aug. 10, 2021).

Critically, agentic AI does not diminish the lawyer’s supervisory role. To the contrary, it heightens the importance of thoughtful workflow design, ongoing oversight, and documentation. Courts evaluating such systems are likely to focus less on nomenclature and more on whether attorneys maintained control, exercised judgment, and could explain their process.

V. Practical Litigation Applications in Complex Matters

Practical adoption of AI in litigation succeeds when it is tied to discrete, outcome-oriented objectives rather than abstract technological ambition. Courts are generally receptive to tools that reduce burden, clarify facts, and streamline dispute resolution—particularly where those tools support early, informed decision-making.

AI-assisted early case assessment allows parties to identify key documents and factual themes at the outset of litigation. Rapid insight into the evidentiary landscape enables counsel to evaluate exposure, defenses, and settlement posture before discovery costs escalate. In many cases, this early clarity reduces the likelihood of protracted discovery disputes driven by incomplete or speculative understandings of the facts.

Deposition preparation illustrates another high-value application. AI tools can aggregate every reference to a witness across emails, memoranda, contracts, and prior testimony, surfacing patterns and inconsistencies that might otherwise remain buried. Counsel can then focus deposition time on resolving factual ambiguities and testing credibility, rather than locating foundational documents.

Chronology building similarly benefits from automation. AI-assisted “living chronologies” update continuously as new data is ingested, helping counsel understand how events unfold over time. This capability is particularly valuable in complex commercial matters involving extended timelines and multiple custodians.

Privilege log review and contractual analysis also benefit from AI-assisted pattern recognition. Tools can flag anomalies—such as non-lawyers included in communications claimed as privileged—or identify non-standard contractual provisions across large document sets. These applications do not replace legal analysis; they accelerate it.

VI. The Enduring Human Element: Ethics, Supervision, and Accountability

Despite increasing automation, the attorney’s role remains central. Rules governing discovery certification, supervision of non-lawyers, confidentiality, and data security apply with full force to AI-assisted workflows.

Certification obligations under Rule 26(g) remain unchanged. Attorneys must certify that discovery responses are complete and correct to the best of their knowledge, formed after a reasonable inquiry. The use of AI does not dilute this responsibility; it reshapes what constitutes a reasonable inquiry in data-intensive cases.

Professional responsibility rules governing supervision apply by analogy to AI systems. Attorneys must understand the capabilities and limitations of the tools they employ and must remain actively involved in decision-making. Courts are unlikely to fault the use of advanced technology where counsel can demonstrate meaningful oversight, quality control, and accountability.

Data security and confidentiality likewise remain critical considerations. AI tools must be deployed in secure environments with appropriate safeguards to protect privileged and sensitive information. When these conditions are met, AI does not undermine ethical obligations; it reinforces them by enabling more accurate, timely, and proportionate discovery.

VII. Conclusion: From Reviewer to Orchestrator

The evolution of AI in litigation reflects a broader shift in the lawyer's role. Attorneys are no longer expected to personally review every document to demonstrate diligence. Instead, they are expected to design, supervise, and defend discovery processes that are reasonable, transparent, and proportional.

As courts continue to refine discovery expectations in an era of expanding data, AI-assisted workflows are likely to become not merely acceptable, but expected in appropriate cases. The critical distinction will not be whether AI was used, but whether its use was thoughtful, defensible, and clearly explained.

The future of litigation belongs not to the fastest reviewers, but to lawyers who can orchestrate facts, technology, and professional judgment into a coherent and defensible strategy.

New York, NY 10019

David Boies
Damien J. Marshall
Andrew Z. Michaelson
BOIES SCHILLER & FLEXNER LLP
575 Lexington Ave.
New York, NY 10022

For defendants Nomura Holding America Inc., Nomura Asset
Acceptance Corp., Nomura Home Equity Loan, Inc., Nomura Credit &
Capital, Inc., Nomura Securities International, Inc., David
Findlay, John McCarthy, John P. Graham, Nathan Gorin, and N.
Dante LaRocca:

Amanda F. Davidoff
Elizabeth A. Cassady
SULLIVAN & CROMWELL LLP
1700 New York Ave., NW, Ste. 700
Washington, DC 20006

Bruce E. Clark
Katherine J. Stoller
SULLIVAN & CROMWELL LLP
125 Broad St.
New York, NY 10004

For defendant RBS Securities Inc.:

Thomas C. Rice
David J. Woll
Andrew T. Frankel
Alan Turner
SIMPSON THACHER & BARTLETT LLP
425 Lexington Ave.
New York, NY 10017

DENISE COTE, District Judge:

Plaintiff Federal Housing Finance Agency ("FHFA"), as
conservator for the Federal National Mortgage Association
("Fannie Mae") and the Federal Home Loan Mortgage Corporation
("Freddie Mac") (together, the Government-Sponsored Enterprises

or "GSEs"), brought these actions against financial institutions involved in the packaging, marketing, and sale of residential mortgage-backed securities ("RMBS") purchased by the GSEs between 2005 and 2007. FHFA has pled claims under Sections 11, 12(a)(2), and 15 of the Securities Act of 1933 ("Securities Act"), as well as the District of Columbia and Virginia Blue Sky laws (together, the "Blue Sky Laws"). FHFA alleges, among other things, that defendants¹ made materially false statements in offering documents for the RMBS. Defendants now move for summary judgment on the ground that FHFA's securities claims are time-barred by the applicable statutes of repose.

BACKGROUND

The GSEs purchased the RMBS at issue in these actions between November 30, 2005 and July 3, 2007. More than three years later, FHFA brought the above-captioned actions. FHFA filed both of these actions on September 2, 2011, within three years of FHFA's appointment as conservator of Fannie Mae and Freddie Mac on September 6, 2008.

¹ The remaining defendants are HSBC North America Holdings Inc. and related entities ("HSBC"), Nomura Holding America Inc. and related entities ("Nomura"), and RBS Securities Inc. ("RBS"). The GSEs purchased the securities at issue between the following dates: with respect to HSBC, from December 20, 2005 to July 3, 2007; with respect to Nomura, from November 30, 2005 to April 30, 2007; and with respect to RBS, from August 31, 2006 to January 31, 2007.

Section 13 of the Securities Act contains a three-year statute of repose that governs claims brought under Sections 11 and 12. 15 U.S.C. § 77m. The Blue Sky Laws contain two- and three-year statutes of repose. D.C. Code § 31-5606.05(f)(1) (three years); Va. Code Ann. § 13.1-522(D) (two years).

In 2008, in the aftermath of the financial crisis, Congress created the FHFA, authorized it to act as conservator for the GSEs, and passed a statute extending FHFA's time to bring any action on their behalf. The Housing and Economic Recovery Act of 2008 ("HERA") creates a new "statute of limitations with regard to any action brought by the [FHFA] as conservator or receiver." 12 U.S.C. § 4617(b)(12)(A). In the case of any "tort claim," "the applicable statute of limitations" is the longer of (1) the three-year period beginning on the date FHFA is appointed as conservator or receiver; (2) the three-year period beginning on the date on which the cause of action accrues; and (3) the period applicable under state law. Id. at § 4617(b)(12). In addition, HERA provides for the revival of tort claims "for which the statute of limitations applicable under State law . . . has expired not more than 5 years before the appointment of the [FHFA]." Id. at § 4617(b)(13)(A). HERA defines "tort claim" to mean "a claim arising from fraud, intentional misconduct resulting in unjust enrichment, or

intentional misconduct resulting in substantial loss to the regulated entity.” Id. at § 4617(b)(13)(B).²

Defendants argue that HERA applies to statutes of limitations to the exclusion of statutes of repose and thus that Section 13’s and the Blue Sky Laws’ two- and three-year statutes of repose bar these actions. Former defendant UBS Americas, Inc. and its affiliates (“UBS”), which later settled the related action brought by FHFA against them, made the same argument to this Court in a motion to dismiss. On June 19, 2012, the Court denied that motion in relevant part, holding that HERA extended statutes of repose as well as statutes of limitations, but certified the issue for interlocutory appeal pursuant to 28 U.S.C. § 1292(b). FHFA v. UBS Americas, Inc., 858 F. Supp. 2d 306 (S.D.N.Y. 2012) (“UBS I”). The June 19 decision was affirmed by the Second Circuit on April 5, 2013. FHFA v. UBS Americas Inc., 712 F.3d 136 (2d Cir. 2013) (“UBS II”). Defendants here raised the same argument in their own motions to dismiss, which were denied in relevant part on November 27, 2012 (Nomura and RBS) and November 28, 2012 (HSBC).

² This provision, subparagraph (13)(B), contains a typographical error, as it refers to “[a] tort claim referred to under clause (i).” The words “tort claim” appear in clause (ii), not clause (i). Immediately preceding subparagraph (13)(B) is subparagraph (13)(A), which refers to “any tort claim described under clause (ii).” Accordingly, it is clear that (13)(B) defines “tort claim” as used in clause (12)(A)(ii).

On June 9, 2014, the Supreme Court issued an opinion in CTS Corp. v. Waldburger, 134 S. Ct. 2175 (2014), holding that a provision in the Comprehensive Environmental Response, Compensation, and Liability Act ("CERCLA") extended certain statutes of limitations but not statutes of repose. Before CTS was decided, the Tenth Circuit held that an extender provision governing actions brought by the National Credit Union Administration ("NCUA") -- a part of the Financial Institutions Reform, Recovery, and Enforcement Act of 1989 ("FIRREA") -- that is nearly identical to the HERA extender provision applied to statutes of repose.³ NCUA v. Nomura Home Equity Loan, Inc., 727 F.3d 1246 (10th Cir. 2013). The Supreme Court vacated and remanded that decision for reconsideration in light of CTS. Nomura Home Equity Loan, Inc. v. NCUA, 134 S. Ct. 2818 (June 16, 2014). On August 19, 2014, the Tenth Circuit reinstated its decision, holding that CTS did not alter its analysis. NCUA v. Nomura Home Equity Loan, Inc., --- F.3d ---, 2014 WL 4069137 (10th Cir. Aug. 19, 2014) ("NCUA").

Defendants brought these motions on June 20, 2014. They were fully submitted on July 18. Defendants contend that CTS so

³ HERA's subsection (b)(12) is identical to FIRREA's subsection (b)(14) -- after changing "Board" to "Agency" and "liquidating agent" to "receiver" -- with one trivial exception: the NCUA statute refers to "the date the claim accrues" where HERA refers to "the date on which the claim accrues." See 12 U.S.C. § 1787(b)(14); 12 U.S.C. § 4617(b)(12).

changes the applicable analysis that this Court is no longer bound by the Second Circuit's opinion. Defendants request summary judgment on this basis and, in the alternative, ask that this Court certify this issue for interlocutory appeal pursuant to 28 U.S.C. § 1292(b). For the reasons that follow, defendants' motions and their requests for certification are denied.

DISCUSSION

The Second Circuit has already decided this question, and the subsequent decision in CTS does not undermine the Second Circuit's ruling for many of the reasons set out in the Tenth Circuit's August 21 Opinion in NCUA, which is incorporated by reference, as well as certain additional reasons given below.

I. Legal Standard

"In construing a statute, we begin with the plain language, giving all undefined terms their ordinary meaning." UBS II, 712 F.3d at 141. Courts are "not to construe each phrase literally or in isolation," but rather to "attempt to ascertain how a reasonable reader would understand the statutory text, considered as a whole." Id. (citation omitted). When statutory text is ambiguous, courts turn to other methods of statutory interpretation, including legislative history. Id.

II. CTS Did Not Disturb the Second Circuit's Holding.

The Second Circuit squarely held that, “[i]n view of the text of the statute and its legislative history . . . it is clear that Congress intended one statute of limitations -- § 4617(b)(12) of HERA -- to apply to all claims brought by FHFA as conservator” and to “supplant[] any other time limitations that otherwise might have applied.” UBS II, 712 F.3d at 143-44. For each of the reasons set out in the Tenth Circuit’s finely written opinion, CTS does not disturb this holding. As the text of FIRREA is identical, in all material respects, to that of HERA, the Tenth Circuit’s textual analyses apply directly.

Among those holdings which this Court adopts, the following are of particular note:

The text and structure of [HERA’s] Extender Statute are fundamentally different from [CERCLA’s]. . . . [B]y establishing all-purpose time limits for any actions [FHFA] may wish to pursue, the Extender Statute displaces all preexisting limits on the time to bring suit, whatever they are called. . . . CERCLA has a completely different structure. Rather than setting its own time limit to bring a [claim], [CERCLA] recognizes that the time limits in state statutes apply.

NCUA, 2014 WL 4069137, at *5 (citation omitted).

[U]nlike [CERCLA], which employs the term “applicable limitations period” to identify the state law time frames modified by the federal commencement date (that is, the specific object of federal preemption), [HERA’s] Extender Statute uses “period applicable under State law” to help construct a new exclusive time framework for [FHFA] actions that replaces all pre-existing time limits (including repose periods). Whether the state period used to

construct this framework is one of limitations or repose has no bearing on whether the new Extender Statute framework itself displaces statutes of repose.

Id. at *9.

In sum, [HERA]'s Extender Statute's surrounding language differs considerably from [CERCLA]'s in that it features the concept of repose, uses the word "period" differently, and lacks a tolling provision. The [Supreme] Court's analysis of the terms "period" and "civil action," as well as the tolling provision in [CERCLA], cannot be extended to [HERA's] Extender Statute because its text and structure are fundamentally different from [CERCLA's].

Id. at *10. The Tenth Circuit's other analyses apply as well for the reasons below.

A. Legislative History and Purpose

The Tenth Circuit emphasized that FIRREA's legislative history and purpose strongly support the court's holding, and contrast sharply with CERCLA's. NCUA, 2014 WL 4069137, at *11-13. The same is true for HERA.

In CTS, the Supreme Court relied, in part, on a report commissioned by Congress that recommended changes to state tort law including the discovery rule enacted in CERCLA's extender provision. This report "acknowledged that statutes of repose were not equivalent to statutes of limitations and that a recommendation to pre-empt the latter did not necessarily include the former." CTS, 134 S. Ct. at 2186. By contrast, the legislative history of HERA strongly confirms that its

limitations provision displaces all previously applicable timeliness provisions.

The Second Circuit explained that

Congress enacted HERA and created FHFA in response to the housing and economic crisis, precisely because it wanted to address the dire financial condition of Fannie Mae and Freddie Mac. As HERA makes clear, Congress intended FHFA to take action to “collect all obligations and money due” to the GSEs, to restore them to a “sound and solvent condition.”

UBS II, 712 F.3d at 142 (quoting 12 U.S.C. § 4617(b)(2)(B)(ii), (D)). With HERA, Congress created the FHFA and vested it with investigatory powers, like the subpoena power, to enable it to suss out the GSEs’ claims. HERA was designed “to give FHFA the time to investigate and develop potential claims on behalf of the GSEs.” Id. As the Second Circuit noted, “Congress obviously realized that it would take time for this new agency to mobilize and consider whether it wished to bring any claims and, if so, where and how to do so.” Id. Accordingly, HERA created a new statute of limitations running, at the earliest, from the appointment of FHFA as conservator, that “supplants any other time limitations that otherwise might have applied” to FHFA’s claims. Id. at 143-44.

B. Statutory Context

The Tenth Circuit considered the use of the phrase “statute of limitations” or “statute of limitation” elsewhere in FIRREA. The court noted that the phrase is used in provisions setting

deadlines for appealing NCUA's denial of a claim that do not allow for accrual or tolling, which is indicative of a broad use of the phrase "statute of limitations" encompassing statutes of repose. NCUA, 2014 WL 4069137 at *10.

HERA's Section 4617 includes two similar provisions setting deadlines for appeal of FHFA's denial of certain claims. See 12 U.S.C. § 4617(a)(6)(B), (8)(D). Both provisions are styled "[s]tatute of limitations," despite the fact that they do not allow for accrual or tolling. Id. "[I]t is a normal rule of statutory construction that identical words used in different parts of the same act are intended to have the same meaning." Taniguchi v. Kan Pac. Saipan, Ltd., 132 S. Ct. 1997, 2004-05 (2012) (citation omitted). Accordingly, these provisions further support a broad construction of "statute of limitations" to encompass statutes of repose.

III. Additional Arguments

Defendants raise several additional arguments not treated in the Tenth Circuit's Opinion. They are addressed in turn.

A. HERA's Revival Provision Refers to a Singular Period.

Defendants argue that the phrase "statute of limitations" must be read narrowly in the HERA provision reviving expired tort claims, subparagraph (13)(A), and that this narrow reading should therefore apply throughout the statute. Defendants misconstrue this provision. It reads:

In the case of any tort claim described [above] for which the statute of limitations applicable under State law with respect to such claim has expired not more than 5 years before the appointment of the [FHFA] as conservator or receiver, the [FHFA] may bring an action as conservator or receiver on such claim without regard to the expiration of the statute of limitations applicable under State law.

12 U.S.C. § 4617(b) (13) (A).

Defendants contend that reading "statute of limitations" broadly here to encompass statutes of repose "would not make sense where the five-year rule dictated a different result for the statute of limitations and the statute of repose." Yet this provision measures five years from the date the limitations period "expired" -- i.e., when the claim became untimely -- which is a single date, whether or not repose periods are included.

Consider, for example, the Securities Act, the model for many states' Blue Sky laws. Section 13 of the Securities Act bars any action brought more than (a) one year after discovery of the untrue statement or omission or (b) three years after the offering or sale of the relevant security. 15 U.S.C. § 77m. Section 13's limitations period "expires" when either the one-year period or the three-year period runs, whichever runs first. It would be odd, indeed, to contend that this period had not "expired" more than three years after the offering or sale, simply because the one-year period had not yet run. It is

defendants' interpretation of the revival provision that is unreasonable, as defendants would have HERA resuscitate claims five years after the discovery-based period had run, but leave untouched claims barred by a two- or three-year repose period. That result would be wholly out of keeping with HERA's structure and purpose.

B. Date of Passage

Defendants next argue that "the distinction between statutes of limitations and repose was clear by the time of HERA's passage." The Supreme Court did note in CTS that the "more precise" usage of "statute of limitations," in distinction to a statute of repose, is "now predominant." CTS, 134 S. Ct. at 2186. But, in the immediately preceding paragraph, the Court "acknowledged that the term 'statute of limitations' is sometimes used in a less formal way" to "refer to any provision restricting the time in which a plaintiff must bring suit." Id. at 2185. In particular, the Court recognized that "Congress has used the term 'statute of limitations' when enacting statutes of repose," and cited in support a provision of the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 -- which was passed two years after HERA. Id. (citing 15 U.S.C. § 78u-6(h)(1)(B)(iii)(I)(aa)). As in CTS, the fact that HERA employs the term "statute of limitations" is "instructive, but it is not dispositive." Id.

C. Securities Claim Is Not a "Tort Claim."

Defendants next contend that HERA applies only to state contract and tort claims, not to statutory claims or federal claims. As an initial matter, the Second Circuit squarely rejected this argument in UBS. "Giving the words of § 4617(b)(12) their plain meaning, and considering the provision as a whole, [the Second Circuit] conclude[d] that a reasonable reader could only understand it to apply to both the federal and state claims in this case." UBS II, 712 F.3d at 142. CTS did not address this issue, and defendants have offered no reason to believe the law in this circuit has changed subsequently.

Indeed, defendants' argument is rebutted by HERA's plain language, as HERA defines "tort claim" to mean "a claim arising from fraud, intentional misconduct resulting in unjust enrichment, or intentional misconduct resulting in substantial loss to the regulated entity." 12 U.S.C. § 4617(b)(13)(B). FHFA's securities fraud claims here are alleged to "aris[e] from fraud . . . or intentional misconduct resulting in substantial loss" to the GSEs, and thus are easily encompassed by this definition. Even if this definition did not apply,⁴ HERA states expressly that it is creating "the applicable statute of

⁴ As noted above, due to a typographical error in the statute, HERA defines "tort claim" as used in "clause (i)" rather than "clause (ii)," despite the fact that "tort claim" only appears in clause (ii).

limitations with regard to any action brought by the Agency as conservator or receiver.” Id. at § 4617(b)(12)(A) (emphasis added).

And defendants are wrong to contend that the reference to otherwise applicable state law periods somehow removes federal claims from HERA’s scope. The fact that the limitations period created for FHFA’s “tort claims” is “the period applicable under State law” in certain circumstances -- where that period extends beyond three years after accrual or appointment of FHFA as conservator -- does not indicate that the only claims covered are state law claims. Where there is no “period applicable under State law” for a “tort claim,” the limitations period is simply three years from accrual or appointment. Again, HERA expressly applies to “any action” brought by FHFA as conservator. Id. at § 4617(b)(12)(A) (emphasis added).

D. Repeal by Implication

Defendants further argue that HERA should not be read to “impliedly repeal” Section 13’s statute of repose, citing cases that predate the Second Circuit’s considered opinion in UBS II holding that HERA “supplants any other time limitations that otherwise might have applied,” including Section 13’s. UBS II, 712 F.3d at 143-44. This Court expressly rejected defendants’ argument in UBS I, 858 F. Supp. 2d at 317 n.8. Defendants offer

no reason why the Court should revisit that decision or second-guess that of the Second Circuit. As explained in UBS I:

Section 13 continues to apply with full force to the vast majority of litigants; HERA creates an exception for a single, privileged plaintiff -- FHFA. Moreover, because, as explained above, HERA's reference to the "statute of limitations" encompasses not only the narrower use of the term advocated by defendants but also what defendants refer to as "statutes of repose," HERA no more impliedly repealed the latter than it did the former. And even defendants agree that, to the extent it applies to federal claims, HERA constitutes a valid extension of Section 13's one-year limitation period.

Id.

E. Presumption Against Pre-emption of State Law

Defendants also argue that a presumption against the pre-emption of state law should apply, citing to a part of Justice Kennedy's opinion in CTS joined by two other Justices. Yet, this presumption is only effective "when the text of a pre-emption clause is susceptible of more than one plausible reading." CTS, 134 S. Ct. at 2188 (citation omitted). For the reasons stated above, that is not the case here.

Moreover, HERA is quite different from CERCLA in this respect. Courts are to "assume[] that the historic police powers of the States [a]re not to be superseded by the Federal Act unless that was the clear and manifest purpose of Congress." Id. (citation omitted). The presumption against pre-emption is strongest "when Congress legislates in an area traditionally

governed by the States' police powers." Id. In CTS, there was "no question that States possess the traditional authority to provide tort remedies to their citizens as they see fit." Id. (citation omitted). Here, by contrast, "[p]olicing fraud against federal agencies is hardly a field which the States have traditionally occupied." Buckman Co. v. Plaintiffs' Legal Comm., 531 U.S. 341, 347 (2001) (citation omitted). Thus, any such presumption applied here would be weak.

IV. Certification for Interlocutory Appeal

The standard for certification is well established.

Section 1292(b) provides, in relevant part, that

[w]hen a district judge, in making in a civil action an order not otherwise appealable under this section, shall be of the opinion that such order involves a controlling question of law as to which there is substantial ground for difference of opinion and that an immediate appeal from the order may materially advance the ultimate termination of the litigation, he shall so state in writing in such order. The Court of Appeals which would have jurisdiction of an appeal of such action may thereupon, in its discretion, permit an appeal to be taken from such order, if application is made to it within ten days after the entry of the order.

28 U.S.C. § 1292(b); see Casey v. Long Island R.R. Co., 406 F.3d 142, 146 (2d Cir. 2005) (noting that Section 1292(b) "imposes both procedural and substantive requirements on a would-be appellant").

The Second Circuit has emphasized that Section 1292(b) certification should be "strictly limited because only

exceptional circumstances will justify a departure from the basic policy of postponing appellate review until after the entry of a final judgment.” Flor v. BOT Fin. Corp., 79 F.3d 281, 284 (2d Cir. 1996) (per curiam) (citation omitted).

Certification is thus appropriate only in the narrow class of cases in which “an intermediate appeal may avoid protracted litigation.” Koehler v. Bank of Bermuda Ltd., 101 F.3d 863, 866 (2d Cir. 1996).

Certification is inappropriate here. At the earliest stages of this massive litigation, this Court certified this very question to the Second Circuit, which issued an opinion that squarely addressed it. For the reasons stated by the Tenth Circuit in NCUA and those given above, it is clear that CTS does not disturb that decision. Accordingly, there is no “substantial ground for difference of opinion.” 28 U.S.C. § 1292(b). Review of the decisions issued recently in the Western District of Texas does not alter that judgment. See FDIC v. Merrill Lynch, Pierce, Fenner & Smith, Inc., 2014 WL 4161561 (W.D. Tex. Aug. 18, 2014); FDIC v. Goldman, Sachs & Co., 2014 WL 4161567 (W.D. Tex. Aug. 18, 2014).

Nor would interlocutory appeal “materially advance the ultimate termination of the litigation.” 28 U.S.C. § 1292(b). Trial in the HSBC action begins on September 29; trial in Nomura is scheduled to begin in several months. The parties will soon


be able to appeal this issue, together with all other issues, following a final judgment. The most efficient way to reach the ultimate termination of this litigation is to try these cases.

CONCLUSION

Defendants' June 20, 2014 motions for summary judgment concerning the applicable statutes of repose and requests for Section 1292(b) certification on this issue are denied.

SO ORDERED:

Dated: New York, New York
August 28, 2014



DENISE COTE
United States District Judge

**IN THE UNITED STATES DISTRICT COURT
FOR THE WESTERN DISTRICT OF PENNSYLVANIA**

	:	
	:	Master Docket
IN RE DIISOCYANATES ANTITRUST LITIGATION	:	Misc. No. 18-1001
	:	
	:	MDL No. 2862
<i>This document relates to: All Actions</i>	:	
	:	

Hon. James C. Francis IV (Ret.)
E-Discovery Special Master

Report and Recommendation

This multidistrict litigation involves allegations that the defendants conspired to reduce supply and increase prices for methylene diphenyl diisocyanate (“MDI”) and toluene diisocyanate (“TDI”), chemicals used in the manufacture of polyurethane foam and thermoplastic polyurethanes. The plaintiffs have moved pursuant to Rules 26 and 37 of the Federal Rules of Civil Procedure to compel defendants BASF Corporation, Covestro LLC, the Dow Chemical Company, Huntsman Corporation, and Wanhua Chemical (America) Co., Ltd. (collectively, the “defendants”)¹ to use certain search terms and technology-assisted review methodologies to identify electronically stored information (“ESI”) for production in response to the plaintiffs’ discovery requests. (ECF 455). These defendants have opposed the plaintiffs’ motion and have cross-moved pursuant to Rule 26(c) for entry of a protective order permitting the defendants to use the search terms and TAR methodologies that they have proposed to search for and review ESI. (ECF 470).

To facilitate resolution, the Court appointed me as E-Discovery Special Master pursuant to Rule 53 to submit a report and recommendation with respect to these motions. (ECF 504). I

¹ This group does not include additional defendants who are not party to these motions.

convened a preliminary conference, conducted a hearing in which technical presentations were provided by the parties' respective experts, and heard oral argument on the motions. In addition, the parties engaged in mediation in an attempt to resolve the issues presented by the motions, but those efforts were unsuccessful. Accordingly, the motions are ripe for determination. I will describe basic aspects of technology-assisted review, set forth the parties' competing methodologies, discuss the governing legal principles, and then apply those principles to the issues in dispute.

As will be seen, the defendants' motion should be denied because their methodologies are not reasonable in certain critical respects. At the same time, the plaintiffs' motion should also be denied because the Court should not foreclose the defendants from choosing alternative methodologies as long as they are reasonable.

Technology-Assisted Review

Technology-assisted review, or "TAR,"

is a process for ranking – from most to least likely to be responsive – or for classifying – as responsive or nonresponsive – a document collection, using computer software that learns to distinguish between responsive and non-responsive documents based on coding decisions made by one or more knowledgeable reviewers on a subset of the document collection. The software then applies what it has learned to the remaining documents in the collection.

(Declaration of Maura R. Grossman dated March 25, 2021 (ECF 459) ("Grossman Decl."), ¶ 9); see also *The Sedona Conference Glossary: E-Discovery & Digital Information Management (Fourth Edition)*, 15 Sedona Conf. J. 305, 357 (2014) (defining TAR as "[a] process for prioritizing or coding a collection of [ESI] 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"). TAR is often referred to as "predictive coding" because the software predicts the classification that a human reviewer would ultimately

assign to a document. TAR software is also variously referred to as the TAR tool, classifier, or algorithm.

It is useful to distinguish between two broad categories of TAR technology. Initially, most TAR tools relied on “seed sets” of documents to train the software. Once the software had reached a point where it would not benefit materially from additional training, it would be applied to the set of unreviewed documents and score each document for likely responsiveness. (Declaration of Daniel L. Regard II dated April 9, 2021 (ECF 471-2) (“Regard Decl.”), ¶ 26(a)); *see also* Bolch Judicial Inst. & Duke Law, Technology Assisted Review (TAR) Guidelines 4 (Jan. 2019) (hereinafter “Duke TAR Guidelines”), *available at* <https://judicialstudies.duke.edu/wp-content/uploads/2019/02/TAR-Guidelines-Final-1.pdf>, at 4; *Lawson v. Spirit AeroSystems, Inc.*, No. 18-1100, 2020 WL 6343292, at *3 (D. Kan. Oct. 29, 2020). This methodology is referred to as TAR 1.0.

By contrast,

[i]n TAR 2.0, a more recent iteration of TAR technology, a classifier is continuously trained in near real-time using the documents that are coded by reviewers. This workflow is also called Continuous Active Learning (“CAL”). Training begins immediately and the classifier improves over time as more documents are reviewed. As with TAR 1.0, the TAR 2.0 classifier assigns a score to each document predicting the likelihood that it is responsive, but unlike TAR 1.0 the classifier adjusts the scores throughout the document review as it learns more. Typically, a TAR 2.0 system prioritizes documents with the highest scores (i.e., the unreviewed documents that TAR predicts are most likely to be responsive) for review. This means that once the TAR classifier has learned enough to score documents, most documents reviewed in the early stage of the review are ranked highly responsive, and there are more of them, and as the review continues, fewer and fewer responsive documents are indicated, and their ranking is lower and lower.

(Regard Decl., ¶ 26(b)(i), (ii)); *see also* Duke TAR Guidelines at 5; *Lawson*, 2020 WL 6343292, at *3; *In re Valsartan, Losartan, and Irbesartan Products Liability Litigation*, 337 F.R.D. 610, 614 (D.N.J. 2020).

In measuring the effectiveness of search and review, two metrics are important.

One is referred to as *recall*, which is a measure of *completeness*, reflected by the proportion (i.e., percent) of responsive documents in a collection that have been found through a search or review process, out of all possible responsive documents in the collection. . . . The other metric is referred to as *precision*, which is a measure of *accuracy*, or the proportion (i.e., percent) of the documents identified by the search or review process that are actually responsive. . . . High recall suggests that substantially all responsive documents have been found; high precision suggests that primarily responsive documents have been found.

(Grossman Decl., ¶ 10); *see also Lawson v. Spirit AeroSystems, Inc.*, No. 18-1100, 2020 WL 1813395, at *7 (D. Kan. April 9, 2020). In addition, another metric relevant to this case is prevalence, also referred to as richness. This is the estimated proportion of responsive documents in a collection at the outset. For example, if there were 10 responsive documents in a set of 100 documents, the richness would be 10%. (Grossman Decl., ¶ 10).

Defendants' Proposed Methodology

In the aggregate, the defendants have collected more than 10 million documents in response to the plaintiffs' discovery demands. (Memorandum of Law in Support of Defendants' Joint Motion for a Protective Order Regarding Certain Search Terms and TAR Methodologies (ECF 471) ("Def. Memo.") at 1). Although there are minor differences among the methodologies suggested by each defendant, each fundamentally involves a five-step process for sifting through this information:

1. The defendant would define the "TAR Review Set" by applying agreed search terms to its data collection and removing items not suitable for TAR (for example, video images or audio files).

2. The defendant would then take a random sample from the TAR Review Set and review it in order to obtain an estimate of the richness of the TAR Review Set, that is, the prevalence of responsive documents. The sample would be designed to achieve a 95% confidence level with a

margin of error of 2%. For example, if the reviewers identified 72 responsive documents in a sample of 2,400 documents randomly selected from the TAR Review Set, it would be estimated that the richness of the entire TAR Review Set was 3%.

3. Next, the defendant would train the TAR predictive tool by “showing” it responsive (and, depending on the tool, non-responsive) documents. The tool would then score the documents within the TAR Review Set according to the likelihood that they are responsive. Because each defendant has indicated that it would use a continuous active learning tool, the determinations with respect to the documents reviewed would then be “fed back” to the classifier, refining its training and allowing it to rescore the documents remaining to be reviewed.

4. The defendant would continue the review process until it estimated that it had achieved a recall rate of 70%. This would be based on a calculation made from the previous estimation of richness. Thus, if the TAR Review Set were 100,000 documents, and the richness estimate was 3%, it would be expected that there would be approximately 3,000 responsive documents to be found. The defendant would pause its review when 70%, or 2,100 responsive documents, had been identified.

5. Finally, the defendant would validate its review by using an elusion test to determine if the 70% recall target had in fact been achieved. This would first involve calculating an estimate of the number of responsive documents that remained in the TAR Review Set and had not been reviewed (often referred to as the “null set”). See *City of Rockford v. Mallinckrodt ARD, Inc.*, 326 F.R.D. 489, 493-94 (N.D. Ill. 2018) (describing sampling of null set in context of search terms). As with the initial richness estimate, this would be done by drawing a sample from the null set with a confidence level of 95% and a margin of error of 2%. Reviewers would then analyze this sample to determine the percentage of responsive documents and extrapolate from

that the estimated number of responsive documents remaining in the null set. For example, if 70,000 documents had been reviewed from the corpus of 100,000 documents, this would leave 30,000 unreviewed documents in the null set. If a sample from that null set showed that 5% of those documents were responsive, it would be inferred that approximately 1,500 responsive documents (i.e., 30,000 x .05) remained in the null set had never been reviewed and had therefore “eluded” detection. From this, the defendant would then calculate the recall rate by dividing the number of responsive documents identified in the review (“R”) by the sum of the identified responsive documents (“R”) plus the estimate of eluded documents (“E”):

$$\text{Recall (\%)} = \frac{\text{Responsive (\#R)}}{\text{Responsive (\#R) + Eluded (\#E)}}$$

In the example above, the number of responsive documents found was 2,100, and the number of eluded documents was 1,500. Therefore, the recall rate is:

$$\text{Recall (\%)} = \frac{2,100}{2,100 + 1,500} = 58\%$$

Since, in this example, the estimated recall rate fell below 70%, the defendant would resume its review process, coding additional batches of documents offered up by the TAR tool, until the 70% threshold had been reached, at which point it would halt its review. (Regard Decl., ¶¶ 31, 35).²

² Although the description of the defendants’ methodology, including the formulas, is derived from the declaration of their primary expert, Daniel L. Regard II, I have provided the numerical example for clarity, and any arithmetic errors are my own.

Plaintiffs' Proposed Methodology

The plaintiffs' proposal focuses on two aspects of the TAR workflow: the TAR review stopping criteria and the validation protocol.

The stopping criteria define the point at which a defendant pauses the review process. As discussed above, the defendants would do so when the number of responsive documents identified surpasses the 70% recall estimate, as indicated by reference to the estimated richness of the TAR Review Set. The plaintiffs' proposal, by contrast, does not rely on an ex ante estimate of the richness of the entire TAR Review Set, but instead depends both on the richness of the last few batches of documents reviewed as well as the content of any responsive documents found in those batches.

The TAR process for each rolling production will continue until Defendants can reasonably conclude that further review is unlikely to yield additional responsive documents with sufficient quantity or materiality to justify continuing. This will not occur before the last two batches of documents identified by TAR and reviewed by humans contains no more than five to ten present (5%-10%) responsive documents, and none of the responsive documents is novel and/or more than marginally relevant.

(Plaintiffs' Proposed Stopping Criterion and Validation Process for Defendants' Application of Technology Assisted Review ("Pl. Proposal") (ECF 460-3) at ECF 2). To make this work, the plaintiffs' proposal requires each defendant to identify which responsive documents came from the last two batches and to reveal the number of documents from those batches deemed responsive but withheld because of privilege. If, on the basis of this information, the parties disagree whether the defendant chose an appropriate stopping point, the proposal requires them to meet and confer and, if no accommodation is reached, to present the dispute to the court. (Pl. Proposal at ECF 2).

The plaintiffs' proposed validation protocol is complex, and is triggered only when, on the basis of the stopping criteria, the defendants "reasonably believe that they have produced or identified for production substantially all responsive non-privileged documents." (Pl. Proposal at ECF 2).

The first step of the validation is the partition of the Tar Review Set into four Subcollections, from each of which a sample would be drawn as follows:

- Subcollection 1: Documents identified by the review as responsive, including privileged documents, from which a random sample of 500 documents would be drawn.
- Subcollection 2: Documents coded as non-responsive by a human reviewer, from which a random sample of 1,500 documents would be drawn.
- Subcollection 3: Documents excluded from review because TAR assigned them a low probability of responsiveness, from which a random sample of 1,500 documents would be drawn.
- Subcollection 4: Documents excluded from review as a result of the application of search terms prior to the application of TAR, from which a random sample of 1,500 documents would be drawn.

(Pl. Proposal at ECF 3).

Next, each of the subsamples would be combined into a single Validation Sample of 5,000 documents that would be presented to one or more Subject Matter Experts ("SMEs") who would, without knowledge of any prior coding determinations or of the subcollection from which any document was drawn, code each document as responsive or non-responsive. (Pl. Proposal at ECF 3). On the basis of the SME(s)' determinations, the following calculations would then be made:

- Number of responsive documents found \approx size of Subcollection 1 x number of responsive docs in the sample from Subcollection 1 \div 500.
- Number of responsive documents coded incorrectly \approx size of Subcollection 2 x number of responsive documents in the sample from Subcollection 2 \div 1,500.

- Number of responsive documents excluded by TAR and not reviewed \approx size of Subcollection 3 x number of responsive documents in Subcollection 3 \div 1,500.
- Number of responsive documents in the set of documents excluded by search terms \approx size of Subcollection 4 x number of responsive documents from the sample of Subcollection 4 \div 1,500.

(Pl. Proposal (Appendix A) at ECF 5). These defined terms would then be combined to calculate estimated recall as follows:

Estimated recall \approx Number of responsive documents found \div (number of responsive documents found + number of responsive documents coded incorrectly + number of responsive documents excluded by TAR and not reviewed + number of responsive documents in the set of documents excluded by search terms).

(Pl. Proposal (Appendix A) at ECF 5). Under the plaintiffs' proposal, while the "estimate of recall should be computed to inform the decision-making process" with respect to when a defendant may stop reviewing, "the absolute number in its own right shall not be dispositive of whether or not a review is substantially complete." (Pl. Proposal (Appendix A) at ECF 5). Rather, the parties would also consider the novelty and materiality of documents now deemed responsive that had previously been coded as non-responsive or had been excluded from review by either search terms or the TAR tool. (Pl. Proposal (Appendix A) at ECF 5).

At this point it is useful to highlight the differences between the defendants' and the plaintiffs' recall estimation processes, an issue that will be revisited later. In broad terms, the formulas are similar in that each is the ratio between the number of responsive documents found and the sum of the number of responsive documents plus the number of responsive documents that were missed. There are, however, important differences. The second term in the denominator of the plaintiffs' formula – the number of responsive documents coded incorrectly – does not appear in the defendants' calculations. Thus, the plaintiffs attempt to account for reviewer error, whereas the defendants do not. Furthermore, the fourth term in denominator of

the plaintiffs' formula – the number of responsive documents in the set of documents excluded by search terms – likewise is omitted from the defendants' formula. Accordingly, while the plaintiffs would calculate recall for both the search term and the TAR segments of the review, the defendants would calculate it only for the TAR portion.³

Governing Legal Principles

Rule 26(g) requires that a party responding to discovery requests must conduct “a reasonable inquiry.” Particularly in the realm of electronic discovery, where the volumes of data can be vast and the costs of locating every responsive document can be enormous, this obligation does not require perfection. *See Lawson v. Spirit AeroSystems, Inc.*, No. 18-1100, 2020 WL 1813395, at *7 (D. Kan. April 9, 2020) (“[W]hile parties must impose a reasonable construction on discovery requests and conduct a reasonable search when responding to the requests, the Federal Rules do not demand perfection.” (citation omitted)); *Winfield v. City of New York*, No. 15-CV-05236, 2017 WL 5664852, at *9 (S.D.N.Y. Nov. 27, 2017) (“[P]erfection in ESI discovery is not required; rather, a producing party must take reasonable steps to identify and produce relevant documents”); *Hyles v. New York City*, No. 10 Civ. 3119, 2016 WL 4077114, at *3 (S.D.N.Y. Aug. 1, 2016) (“[T]he standard is not perfection, or using the “best” tool, but

³ At oral argument, the defendants suggested for the first time that they, in fact, intended to validate their search by calculating an estimate of recall that took into account documents excluded both by search terms and by TAR (Transcript of Oral Argument dated Aug. 2, 2021 (“Tr.”) at 68-69). That would be a marked change in position. The defendants’ stated validation process defines the “TAR Review Set” as the documents identified after application of search terms and then calculates recall on the basis of documents remaining in the TAR Review Set that had not been subject to review. (Regard Decl., ¶ 31). Moreover, the defendants criticized the plaintiffs’ proposed validation method precisely because it incorporates the search term phase. (Defendants’ Joint Memorandum of Law in Opposition to Plaintiffs’ Motion to Compel Defendants to Use Certain Search Terms and TAR Methodologies (“Def. Opp. Memo.”) (ECF 492) at 7; Declaration of Stephen Shebest in Support of Defendants’ Motion for a Protective Order dated April 8, 2021 (“Shebest Decl.”) (ECF 471-3), ¶ 18).

whether the search results are reasonable and proportional.” (citation omitted)); *Enslin v. Coca-Cola Co.*, No. 2:14-cv-06476, 2016 WL 7042206, at *3 (E.D. Pa. June 8, 2016).

As applied to the complexities of TAR, the principle of reasonableness incorporates an obligation for the producing party to validate its search. While it might be appropriate in a simple case to assume that a properly instructed custodian can conduct a reasonable search of a paper file without the need for a post hoc evaluation of the custodian’s search process, the same cannot be said for a technology-assisted search of hundreds of thousands of documents. Accordingly, courts have required parties to validate their TAR search methodologies. *See Mercedes-Benz Emissions Litigation*, No. 2:16-CV-881, 2020 WL 103975, at *2 (D.N.J. Jan. 9, 2020) (finding that “case law dictates that appropriate validation be utilized to test search results”); *City of Rockford*, 326 F.R.D. at 494 (holding that “validation and quality assurance are fundamental principles to ESI production”). This requirement for validation applies to search terms just as it does to TAR. *See, e.g., In re Seroquel Products Liability Litigation*, 244 F.R.D. 650, 662 (M.D. Fla. 2007) (“[W]hile key word searching is a recognized method to winnow relevant documents from large repositories, use of this technique must be a cooperative and informed process Common sense dictates that sampling and other quality assurance techniques must be employed to meet requirements of completeness.”).

A further relevant legal principle is that “a producing party has the right in the first instance to decide how it will produce its documents.” *In re Valsartan*, 337 F.R.D. at 617; *see also Winfield*, 2017 WL 5664852, at *9; *Hyles*, 2016 WL 4077114, at *2. This general rule does not, however, give carte blanche to a producing party. Rather, “a producing party is best situated to determine its own search and collection methods *so long as they are reasonable.*” *Nichols v. Noom, Inc.*, No. 20-CV-3677, 2021 WL 948646, at *2 (S.D.N.Y. March 11, 2021) (emphasis

supplied). Thus, “ ‘[p]arties cannot be permitted to jeopardize the integrity of the discovery process by engaging in halfhearted and ineffective efforts to identify and produce relevant documents.’ ” *Winfield*, 2017 WL 5664852, at *9 (quoting *HM Electronics, Inc. v. R.F. Technologies, Inc.*, No. 12-cv-2884, 2015 WL 471498, at *12 (S.D. Cal. Aug. 7, 2015), *vacated in part on other grounds*, 171 F. Supp. 3d 1020 (S.D. Cal. 2016)). Moreover, the right of a producing party to choose a general search method – search terms rather than TAR, for example – does not mean that a court must blindly accept all of the specific details of the proffered methodology. For instance, a party that could legitimately to use search terms as its preferred method could not then specify terms that would exclude large swaths of relevant documents. Thus, the decisions holding merely that a court should defer to the producing party’s choice to use a generic search methodology, *see, e.g., In re Viagra (Sildenafil Citrate) Products Liability Litigation*, No. 16-md-02691, 2016 WL 7336411, at *1-2 (N.D. Cal. Oct. 14, 2016) (refusing to compel use of predictive coding); *Hyles*, 2016 WL 4077114 at *3 (finding that “it is not up to the Court, or the requesting party . . . , to force the . . . responding party to use TAR when it prefers to use keyword searching”), provide little guidance in a case like this, where the defendants’ decision to use TAR is not in dispute.

Discovery is also informed by the related, but not identical, concepts of cooperation and transparency. “Technology-assisted review of ESI . . . require[s] an ‘unprecedented degree of transparency and cooperation among counsel.’ ” *Youngevity International Corp. v. Smith*, No. 16-cv-00704, 2019 WL 1542300, at *12 (S.D. Cal. April 9, 2019) (quoting *Progressive Casualty Insurance Co. v. Delaney*, No. 2:11-CV-00678, 2014 WL 3563467, at *10 (D. Nev. July 18, 2014)); *accord In re Valsartan*, 337 F.R.D. at 622. The principle of cooperation is not only established in the case law, it is mandated here both by the Court’s order governing discovery of

ESI in this case (Stipulated Order re: Discovery of Electronically Stored Information dated April 27, 2020 (ECF 313) (“ESI Order”), ¶ 2) and by the Local Rules for the Western District of Pennsylvania (Appendix LCvR 26.2C, Guideline 1.02). However, by itself the requirement to cooperate only goes so far. Cooperation requires good faith efforts to reach agreement on material issues, but it does not require that agreement always be reached. Parties may fulfill their cooperation obligations without resolving all disputes between them.

Transparency transcends cooperation. It does not mean merely that parties must discuss issues concerning the discovery of ESI; it requires that they disclose information sufficient to make those discussions, as well as any court review, meaningful. In this case, the parties agreed, and the Court ordered that “[i]f a party decides to use a computer/technology assisted review (“TAR”) for identifying potentially responsive documents, it shall inform the other parties and, if requested, shall discuss methods and procedures for implementing it.” (ESI Order, ¶ 5(a)(i)). Transparency, then, may go beyond simply identifying the nature of the search methodology to be used. *See Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125, 128 (S.D.N.Y. 2015) (declining to rule on seed set transparency where parties agreed to disclose all non-privileged documents in control sets); *Bridgestone Americas, Inc. v. International Business Machines Corp.*, No. 3:13-1196, 2014 WL 4923014, at *1 (M.D. Tenn. July 22, 2014) (approving party’s agreement to reveal seed set used to train TAR); *In re Actos Products Liability Litigation*, No. 6:11-md-299, 2012 WL 7861249, at *4-5 (W.D. La. July 27, 2012) (noting that experts for each party reviewed and coded TAR seed set pursuant to parties’ ESI protocol). As the court observed in *Winfield*, “[c]ourts are split as to the degree of transparency required by the producing party as to its predictive coding process.” *Winfield*, 2017 WL 5664852, at *10.

The procedural mechanisms that the parties rely on here in seeking the Court's intervention are a motion to compel discovery under Rules 26 and 37 and a motion for a protective order under Rule 26(c). Pursuant to Rule 37(a)(3)(B)(iv), "[a] party seeking discovery may move for an order compelling . . . production . . . if . . . a party fails to produce documents or fails to respond that inspection will be permitted – or fails to permit inspection – as requested under Rule 34." The scope of discovery, in turn, is defined by Rule 26.

Unless otherwise limited by court order, the scope of discovery is as follows: 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, 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. Information within this scope of discovery need not be admissible in evidence to be discoverable.

Fed. R. Civ. P. 26(b)(1). *See generally Southeastern Pennsylvania Transportation Authority v. Orrstown Financial Services, Inc.*, 367 F. Supp. 3d 267, 276 (M.D. Pa. 2019); *Focht v. Nationstar Mortgage, LLC*, No. 3:18-cv-151, 2019 WL 3081625, at *1 (W.D. Pa. July 15, 2019).

Under Rule 26(c), a "court may, for good cause, issue an order to protect a party or person from annoyance, embarrassment, oppression, or undue burden or expense." Fed. R. Civ. P. 26(c)(1). The moving party bears the burden of demonstrating good cause, and must do so by showing that disclosure will result in clearly defined specific and serious injury; broad allegations of harm are insufficient. *Shingara v. Skiles*, 420 F.3d 301, 306 (3d Cir. 2005). In determining whether good cause exists, "the court must balance the need for information against the injury that might result if uncontrolled disclosure is compelled." *Engage Healthcare Communications, LLC v. Intellisphere, LLC*, No. 12-cv-00787, 2016 WL 11680950, at *2 (D.N.J. Nov. 30, 2016) (quotation marks and citation omitted).

I will now address the merits of the parties' motions. Because the defendants' production obligations are at issue, I will begin with their motion.

Defendants' Motion for a Protective Order

A. TAR Methodology

The defendants essentially seek the Court's blessing for their search terms and TAR methodologies in advance of having conducted any review of the bulk of their documents. Assuming that this is an appropriate subject for a protective order in the first place, the motion should be denied because the proposed methodologies contain serious flaws that would preclude the defendants from certifying their discovery responses to be reasonable under Rule 26(g).

First, the defendants' validation protocol is defective because it is limited to the TAR portion of the search and review process. As the defendants recognize, selecting documents for production involves a succession of independent steps, each of which is designed to help identify responsive information and exclude non-responsive data. (Tr. at 21). Initially, documents are collected from the entire universe of a defendant's data, by the identification of custodians and other specific data sources. Next, as in this case, search terms may be applied to filter the information further. The remaining documents become the TAR Review Set to which the TAR algorithm is applied. The documents recommended by the TAR tool as most likely responsive are then analyzed by human reviewers who make a final determination of responsiveness before the documents are produced. At each of these stages except the last, responsive documents may be left behind and never reviewed by humans, and at the final stage, responsive documents may be mistakenly coded as non-responsive.

In this case, the parties have not disputed the defendants' initial collection processes. And they have generally agreed that 70-80% estimated recall is often considered reasonable by the

courts. (Letter of Sarah R. LaFreniere dated March 4, 2021 (ECF 458-8) (“LaFreniere 3/4/21 Ltr.”), attached as Exh. G to Declaration of Sarah R. LaFreniere dated March 25, 2021 (ECF 458), at 1). But what that recall figure is to be derived from is a matter of dispute. The defendants suggest that the parties agreed that 70% recall based on search and review only of the Tar Review Set would be deemed reasonable. (Def. Memo. at 14). Not so. First, any such agreement would have to be predicated on an agreement as to what constitutes the TAR Review Set, something that has not been achieved since the parties have not agreed on the search terms that would define that set. Second, even had there been such an agreement, it is apparent that the plaintiffs anticipated estimating recall based on each step of the process following the initial collection, as illustrated by the stratified sampling method that they proposed which included documents eliminated from consideration by search terms. (LaFreniere 3/4/21 Ltr. at 5-6; *see also* Letter of Hallie E. Spraggins dated Feb. 19, 2021 (ECF 458-6), attached as Exh. E to LaFreniere 3/25/21 Decl., at 4-6 & App. A; Letter of Alden L. Atkins dated March 16, 2021 (ECF 458-9), attached as Exh. H to LaFreniere 3/25/21 Decl., at 6). There was thus no meeting of the minds as to the segments of the search and review process that would be included in calculation of the recall estimate.

In the absence of such agreement, it would be plainly unreasonable to calculate estimated recall for the TAR portion of the process alone. Dr. Maura R. Grossman, the plaintiffs’ expert, provides an example showing why this is the case. Suppose a party collected one million documents, of which 100,000 are responsive. If search terms cull out 600,000 documents, then 400,000 become the TAR Review Set. Suppose further that of the 400,000 documents to which TAR is applied, 70,000 are responsive. That means that the search term portion of the workflow would have left 30,000 responsive documents behind, and the estimated recall for that stage

would be 70% (that is, $70,000 \div (70,000 + 30,000)$). Then assume that the TAR review – the only part of the process that the defendants propose to validate – reached their target of 70% recall because 49,000 responsive documents were identified ($49,000 \div (49,000 + 21,000) = 70\%$). The overall estimated recall for the search term and TAR stages would then be the product of the recall for each stage: 49% ($70\% \times 70\%$, or, calculated as “end-to-end recall, $49,000 \div (49,000 + 51,000)$). (Affidavit of Maura R. Grossman dated March 25, 2021 (ECF 459), ¶ 37). In other words, a defendant would claim a recall rate of 70% when, in fact, it had produced less than half of the responsive documents.

The second flaw in the defendants’ protocol is its failure to take advantage of the specific characteristics of CAL that provide the parties a means for predicting not just the quantity, but also the quality, of likely responsive documents that have not yet been reviewed by humans. When parties have utilized either search terms exclusively or earlier versions of TAR, courts have occasionally required that the results of their searches be validated in part by allowing the requesting party to review a sample of the “null set,” that is, the set of documents deemed non-responsive by the search device and not subjected to human review. *See City of Rockford*, 326 F.R.D. at 493-95 (review of sample of documents excluded by search terms); *Winfield*, 2017 WL 5664852, at *11 (review of sample of documents excluded by TAR). Some courts are understandably reluctant to require follow this course, at least in the absence of a showing of some deficiency in the production, because it seems contrary to the rules of discovery to require disclosure of non-responsive documents.

CAL provides a means for achieving greater confidence in the completeness of a search without the need to reveal non-responsive documents. Because the CAL tool is continuously being retrained and is reranking as yet unreviewed documents in response to what it “learns”

from the coding of documents that have been reviewed by humans, the number of likely responsive documents in each successive batch to which the tool is applied diminishes steadily after the algorithm has been sufficiently trained. As Mr. Regard, the defendants' expert, has attested, "[i]n at TAR review, once the TAR system is trained, a large proportion of the documents batched for review in the early stages are responsive. As the review continues, the proportion of responsive documents declines, often significantly." (Regard Decl., ¶ 19). Thus, by looking at the number of responsive documents in the last couple of batches analyzed by the tool (or, more precisely, by looking at the percentage of responsive documents out of the total number of documents reviewed in each of those batches), a party can evaluate the trend and predict the estimated number of likely responsive documents in the next batch. Perhaps more importantly, a party can look at the content of the responsive documents identified from the last couple of batches to assess their significance and thereby predict whether documents in the next batch would likely be not merely responsive, but important to the litigation. For example, suppose the last batches of documents contain invoices which, although technically responsive, are of little import to the outcome of the case. In that instance, there would be little reason to continue the search. On the other hand, if those batches contain "hot" documents, searching additional batches would be prudent.

The defendants argue that because parties commonly rely on recall as the sole basis for validation, their approach is reasonable. But what constitutes reasonable conduct must necessarily be measured against the available technology. Using validation based exclusively on elusion testing and recall statistics may be reasonable for parties using only search terms or TAR 1.0. But CAL gives the parties a powerful method for evaluating search at the margin, helping

them decide whether further search and review will be proportional. For the defendants to have this tool in their hands but to forgo its use is unreasonable.

Because the defendants' proposed TAR procedures include a validation process that is fundamentally flawed, their motion for a protective order permitting them to utilize those procedures should be denied.

B. Search Terms

The defendants further seek a protective order permitting them to utilize their designated search terms to cull their documents prior to the application of TAR. The parties agreed that search terms may be used for this purpose. (Declaration of Zachary K. Warren dated April 9, 2021 (ECF 472) ("Warren Decl."), Exh. 1; Pl. Memo at 1). They also agreed to a procedure for collaborating on and, if necessary, obtaining a ruling on disputed search terms. (ESI Order, ¶ 5(a)(2)). That process has broken down.

At the point where the parties reached impasse, the plaintiffs' suggested search strings returned between 67% and 87% of the documents collected, depending upon defendant, while the hit rate for the defendants' suggested search strings ranged from 29% to 55%:

	Defendants' March 18 Proposal	Plaintiffs' March 22 Proposal
BASF	55%	81%
Covestro	46%	79%
Dow	44%	87%
Huntsman	38%	75%
Wanhua	29%	67%

(Warren Decl., ¶ 24). The defendants argue that if they are not permitted to use their search terms and are instead required to use the plaintiffs' terms, they will be forced to review large numbers of irrelevant documents at disproportionate cost and burden.

Yet the defendants have done no systematic testing of the proposed terms to determine whether their search strings leave behind significant numbers of responsive documents captured by the plaintiffs' terms. This has two consequences. First, without such testing, the defendants cannot represent that their proposed search terms are substantively reasonable. As the defendants' expert has written,

[t]ypically, in a modern, iterative development process, false negatives are examined to understand why a set of search terms might be missing a known responsive document. This technique, when available, can be used to identify new query strings that will add to the overall *recall* of a set of search terms. This is often used to calibrate search terms and can lead to *recall* improvement.

(Dan Regard & Tom Matzen, *A Re-Examination of Blair & Maron (1985)*, ICAL 2013 DESI V Workshop, Position Paper at 10-11 (June 14, 2013), *available at* <https://docplayer.net/2626527-A-re-examination-of-blair-maron-1985.html>). Second, without testing it is impossible to make a proportionality assessment, since the incremental value of the additional search terms proposed by the plaintiffs remains unknown. As will be discussed further below, the defendants did conduct some testing of their search terms against a portion of their collection: documents that they produced in the context of a Department of Justice investigation. (Declaration of Trisha Jhunjhnuwala dated April 30, 2021 (ECF 496, ECF 497 (sealed)) (“Jhunjhnuwala Decl.”), ¶ 13). But the results of testing a collection of documents already selected as responsive cannot be generalized to the far broader collection to which the defendants would apply their search terms in order to create the TAR Review Set.

Furthermore, the large percentage of documents collected by the plaintiffs' search terms, standing alone, is not meaningful. In discussing search terms that retrieved 80 percent of an office's emails, the District of Columbia Court of Appeals observed that, "far from showing bad faith, that figure may simply indicate that most of the emails actually bear some relevance, or at least include language captured by reasonable search terms." *In re Fannie Mae Securities Litigation*, 552 F.3d 814, 821 (D.C. Cir. 2009). In this regard, it should be kept in mind that the function of search terms in this case is not to identify documents for production or even to select those that will be provided directly to human reviewers; it is to narrow the universe of documents to which TAR will be applied. In this context, precision, which is what the defendants appear to seek, is relatively less important than recall.⁴

The defendants' focus on hit rates without regard for the substance of documents excluded by search terms is particularly problematic given the nature of this case. "The choice of a specific search and retrieval method will be highly dependent on the specific legal context in which it is to be employed. Parties and their counsel must match the use case with the tools and best practices appropriate to address it" *The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery*, 15 Sedona Conf. J. 217, 244 (2014). Here, the plaintiffs' allegations of an antitrust conspiracy suggest that broader search terms may be warranted than in a case where pertinent communications might be less guarded. One specific dispute illustrates this. The defendants contend that the plaintiffs' search terms are overly general: "[t]he term 'in person,' for example, is only relevant as it relates to a meeting with a competitor, so it should be limited to documents that include a competitor's name." (Def. Memo. at 19). But in everyday communications, let alone communications that

⁴ Precision remains significant to the extent that it affects the cost of implementing TAR, as will be discussed below.

might lead to antitrust liability, a person is at least as likely to say, “I’m having an in person meeting with Joe” as they are to say, “I’m having an in person meeting with Covestro.” Yet the defendants’ narrower approach would preclude documents like the hypothetical email about the meeting with Joe from even reaching the stage of TAR review.

The defendants have not only failed to show that their suggested search terms are reasonable, they have also failed to demonstrate that using more expansive terms, even if not those suggested by the plaintiffs, will create an unacceptable burden. Some of this appears to be related to a misunderstanding of the relation between the search terms and the TAR review. For example, in rejecting proposed terms that the defendants consider too generic, they contend that they “should not be required to review every document relating to lunch prices or meeting minutes.” (Def. Memo at 19). But because the defendants are not proposing to do a full manual review following the application of search terms, there is no need to review “every document.” Rather, since the CAL tool ranks documents according to responsiveness, documents relating to lunch prices or meeting minutes that do not also have some other indicator of relevance would soon be ranked near the bottom of the Tar Review Set and would likely never be assessed by a human reviewer.

The defendants’ TAR vendors make somewhat more subtle arguments regarding burden. First, they contend that the more documents that are included in the TAR Review Set, the higher the hosting costs. (Declaration of Tony Reichenberger dated April 9, 2021 (ECF 471-6), ¶ 9; Declaration of Joseph Goodman dated April 9, 2021 (ECF 471-4), ¶ 6); *see also Lawson v. Spirit AeroSystems, Inc.*, No. 18-1100, 2020 WL 3288058, at *19 (June 18, 2020) (finding that “the volume of data subjected to the TAR process materially impacts technology costs such as data processing and hosting”). Second, they argue that because the additional documents included if

the plaintiffs' search terms are used would reduce the richness of the TAR Review Set while also potentially adding novel concepts and issues, it would be harder to train the CAL algorithms. In other words, more human review would be required at the front end before the tools begin returning a high proportion of responsive documents. (Shebest Decl., ¶ 10); Goodman Decl., ¶ 7; Declaration of Xavier P. Diokno dated April 9, 2021 (ECF 471-5), ¶ 26); *see also Lawson*, 2020 WL 3288058, at *19 (finding that “[w]hen ‘the TAR set’s richness is extremely low, human reviewers may have a difficult time training the software on what is relevant, because examples may be scarce or difficult to come by in the TAR set’ ” (quoting Duke TAR Guidelines at 27)).

These points are well taken. However, the defendants' efforts to quantify this burden are less successful. Although hosting costs could presumably be calculated simply on the basis of the volume of additional documents captured by the plaintiffs' search terms, the defendants have not specified those costs. Some defendants have attempted to estimate additional review costs by making the assumption that to achieve 70% recall, approximately 75% of the documents would need to be reviewed. They then project review costs by applying this 75% figure to the number of additional documents collected by the plaintiffs' search terms. (Shebest Decl., ¶¶ 11-13). But the defendants acknowledge that the 75% figure “will vary for each set of documents and each case.” (Shebest Decl., ¶ 11). Moreover, Dr. Grossman vigorously disputes the 75% assumption, pointing to academic studies suggesting that, on average, parties that use CAL review only one to two non-responsive documents for each responsive document. (Supplemental Declaration of Marua R. Grossman dated April 30, 2021, ¶¶ 33-35, 70).

Thus, the defendants have not provided sufficiently reliable cost estimates to justify the wholesale rejection of the plaintiffs' search terms, nor have they shown that their own search terms will capture a reasonable proportion of the responsive documents. Accordingly, the

defendants' motion for a protective order allowing them to use their suggested search terms should be denied.

Plaintiffs' Motion to Compel

That the defendants' motion should be denied does not, however, mean that the plaintiffs' motion to compel should be granted. The principle that the producing party is the master of its methodology is a deterrent to imposing a requesting party's proposed procedures unless it is evident that the producing party is unable to come up with a reasonable alternative. That is not the case here. While the defendants' search term and TAR methodologies are not reasonable, the plaintiffs' proposals go beyond what the law requires in at least some respects.

A. TAR Methodology

The plaintiffs contend that the defendants should be required to continue running their CAL tools until two sequential batches contain less than 5-10% responsive documents and no responsive documents that are novel or more than marginally relevant. As discussed above, when the search process is on the brink of completion, it is critical for the parties to give attention both to the richness of the last batches and to the significance of documents found in those batches. However, it would be arbitrary to impose rigid stopping criteria. For example, if more than 10% of the documents in the last few batches were technically responsive, it might nevertheless be reasonable to stop if, as in the example cited earlier, those responsive documents were all invoices of marginal value to the litigation. Furthermore, to the extent that the plaintiffs suggest that the TAR methodology must identify "substantially all responsive non-privileged documents" in order to be valid, they misstate the legal standard, which requires reasonableness, not perfection.

Similarly, the plaintiffs' validation criteria overreach by incorporating an analysis of the accuracy of the determinations made by human reviewers. There is no doubt that reviewers often come to differing conclusions regarding the responsiveness of the same documents. *See, e.g.,* Marua R. Grossman & Gordon V. Cormack, *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, XVII Rich. J.L. & Tech. 1, 10-13 (2011). And there is also no doubt that the determinations of reviewers are often wrong when judged against the determinations of a subject matter expert, as on a second-level review. *Id.* at 13-14. This is true whether the reviewers are engaged in a linear review of all documents collected or in a review of documents culled by search terms or by TAR. Yet courts have generally not mandated the incorporation of such reviewer error in calculating the recall statistic in any of these scenarios. This may be due, in part, to the fact that by the time recall is being calculated, the ability to correct for reviewer error has long since passed. Responsible producing parties instead monitor reviewer performance on an ongoing basis, so that a reviewer who is making erroneous responsiveness determinations can be retrained or replaced and the determinations they have already made can be audited.

It is true, as the plaintiffs argue, that failure to account for reviewer error will artificially inflate the recall statistic because that statistic will include non-responsive documents incorrectly coded as responsive. (Grossman Decl., ¶ 37). But what is deemed an acceptable recall estimate can be adjusted to take human error into consideration. For example, if 70% recall were considered reasonable if human error were included in the calculation, perhaps 72% or 75% or some other percentage would be reasonable if human error is not taken into account.

Thus, since there are alternative TAR methodologies that the defendants could utilize and that are reasonable, they should not be compelled to adopt the plaintiffs'.

B. Search Terms

Similarly, whatever the flaws in the defendants' search terms, the plaintiffs have failed to demonstrate that their proposed search terms are the only reasonable option. For example, the plaintiffs applied their disputed search terms to the documents that the defendants had produced to the Department of Justice and found that these terms resulted in the identification of many arguably responsive documents. (Pl. Memo. at 2-3 & Exh. A). From this, they argue that use of the defendants' terms would have excluded these documents from the TAR Review Set altogether. But that is not the case. The defendants have demonstrated that, with one exception, every one of the exemplary documents cited by the plaintiffs would have been elicited by the defendants' proposed search terms. (Jhunjhnuwala Decl., ¶¶ 19-20). The single exception is a document that does not appear to be responsive to the plaintiffs' document requests (Jhunjhnuwala Decl., ¶ 21), so its exclusion from the TAR Review Set would not prejudice the plaintiffs.

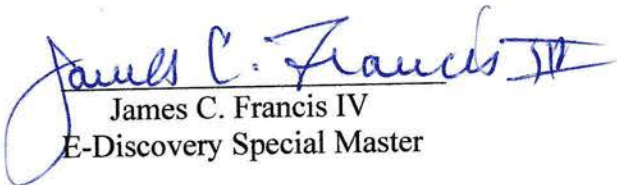
Thus, while the plaintiffs' search terms (or, indeed, the use of no search terms at all) may be a reasonable means of creating the TAR Review Set, there is no basis for precluding the defendants from utilizing a different set of search terms that is also reasonable.⁵ Accordingly, insofar as it relates to search terms, the plaintiffs' motion to compel should be denied.

⁵ The ESI Order provides a mechanism for adjudicating disputes over search terms. (ESI Order, ¶ 5(a)(2)). If need be, the parties can seek a court decision with respect to their disagreements over individual terms.

Conclusion

For the reasons set forth above, I respectfully recommend that Plaintiffs' Motion to Compel Defendants to Use Certain Search Terms and TAR Methodologies (ECF 455) and Defendants' Joint Motion for a Protective Order Regarding Certain Search Terms and TAR Methodologies (ECF 470) each be denied.

Respectfully Submitted,


James C. Francis IV
E-Discovery Special Master

8/23/21
Date

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----X
Monique Da Silva Moore; Maryellen
O'Donohue; Laurie Mayers; Heather
Pierce; and Katherine Wilkinson, on behalf of
themselves and all others similarly situated,

Plaintiffs,

11 Civ. 1279 (ALC) (AJP)

- against -

Opinion & Order

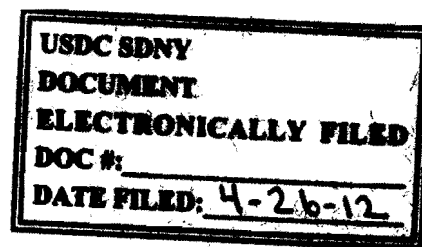
Publicis Groupe SA and MSLGroup,

Defendants.

-----X
ANDREW L. CARTER, JR., United States District Judge:

Before the Court are Plaintiffs' objections, filed pursuant to Rule 72(a) of the Federal Rules of Civil Procedure, to Magistrate Judge Andrew J. Peck's rulings rendered during a February 8, 2012 discovery conference and his February 24, 2012 opinion and order (dkt. no. 96). Judge Peck's rulings and written order discussed, *inter alia*, the use of the predictive coding software, a computer assisted form of review. The Court assumes the parties' familiarity with the facts and the predictive coding method.

Plaintiffs and Defendant MSLGroup entered into a stipulation governing MSLGroup's production of electronically stored information ("ESI protocol"). The substance of the ESI protocol resulted from a series of court conferences and party discussions, but the February 8 conference set the final parameters for the protocol. Judge Peck so-ordered the ESI protocol on February 22, 2012. (Dkt. No. 92.) The last paragraph of the protocol provides that Plaintiffs object to the entire ESI protocol, but signed the agreement because it reflects Judge Peck's discovery rulings and Judge Peck rejected Plaintiffs' version of the ESI protocol.



Plaintiffs object to the February 8 discovery rulings, the ESI protocol, and the February 24 opinion and order, arguing, *inter alia*, that the predictive coding method contemplated in the ESI protocol lacks generally accepted reliability standards, that the use of such method violates Fed. R. Civ. P. 26 and Federal Rules of Evidence 702, that Judge Peck improperly relied on outside documentary evidence in his February 24 opinion and order, that MSLGroup's expert is biased because the use of the predictive coding method will reap financial benefits for the company, that Judge Peck failed to hold an evidentiary hearing, and that he adopted MSLGroup's version of the ESI protocol on an insufficient record. Plaintiffs request that the Court overturn the Magistrate Judge's rulings because they are erroneous and contrary to law.¹ Plaintiffs also submitted a letter requesting that Judge Peck recuse himself from the action, which Judge Peck denied on April 2, 2012, but allowed them to file a formal motion. Plaintiffs filed their recusal motion on April 13, 2012, incorporating similar arguments made in their Rule 72(a) objections. (Dkt. No. 169.)

Rule 72(a) provides that for nondispositive orders issued by a magistrate judge, "[t]he district judge in the case must consider timely objections and modify or set aside any part of the order that is clearly erroneous or is contrary to law." Fed. R. Civ. P. 72(a); *see also* 28 U.S.C. § 636(b)(1)(A). "Under this highly deferential standard of review, magistrates are afforded broad discretion in resolving [non-dispositive] disputes and reversal is appropriate only if their discretion is abused." *AMBAC Fin. Servs., LLC v. Bay Area Toll Auth.*, No. 09 Civ. 7062, 2010 WL 4892678, at *2 (S.D.N.Y. Nov. 30, 2010) (citation omitted). A magistrate judge's ruling is considered "contrary to law" when it "fails to apply or misapplies relevant statutes, case law, or rules of procedure." *In re Comverse Tech., Inc. Sec. Litig.*, No. 06 Civ. 1825, 2007 WL 680779,

¹ Plaintiffs filed their objections to Judge Peck's February 8 rulings on February 22, 2012 and Judge Peck issued his opinion and order on February 24, 2012. (Dkt. Nos. 93-96.) Judge Peck addressed some of Plaintiffs' objections in his opinion and order. Plaintiffs had an opportunity to respond to the written order and they submitted their reply brief on March 19, 2012. (Dkt. Nos. 123-25.)

at *2 (E.D.N.Y. Mar. 2, 2007). “The reviewing court must be left with the definite and firm conviction that a mistake has been committed to overturn the magistrate judge's resolution of a nondispositive matter.” *AMBAC Fin. Servs.*, 2010 WL 4892678, at *2 (citation and internal quotation omitted). “Matters concerning discovery generally are considered ‘nondispositive’ of the litigation.” *Thomas E. Hoar, Inc. v. Sara Lee Corp.*, 900 F.2d 522, 525 (2d Cir. 1990).

Mindful of this highly deferential standard of review, the Court adopts Judge Peck’s rulings because they are well reasoned and they consider the potential advantages and pitfalls of the predictive coding software. The Court has thoroughly reviewed the ESI protocol along with the parties’ submissions.² At the outset, the Court notes that Plaintiffs and Judge Peck disagree about the scope of Plaintiffs’ acquiescence concerning the use of the method. Judge Peck’s written order states that Plaintiffs have consented to its use, (Opinion and Order at 17 (“The decision to allow computer-assisted review in this case was relatively easy – the parties agreed to its use (although disagreed about how best to implement such review.”))), while Plaintiffs argue that Judge Peck’s order mischaracterizes their position (Pl. Reply, dated March 19, 2012, at 4-5). Nevertheless, the confusion is immaterial because the ESI protocol contains standards for measuring the reliability of the process and the protocol builds in levels of participation by Plaintiffs. It provides that the search methods will be carefully crafted and tested for quality assurance, with Plaintiffs participating in their implementation. For example, Plaintiffs’ counsel may provide keywords and review the documents and the issue coding before the production is made. If there is a concern with the relevance of the culled documents, the parties may raise the issue before Judge Peck before the final production. Further, upon the receipt of the production, if Plaintiffs determine that they are missing relevant documents, they may revisit the issue of

² The predictive coding method is provided in pages 10-18 of the ESI protocol. (Dkt. No. 92.)

whether the software is the best method. At this stage, there is insufficient evidence to conclude that the use of the predictive coding software will deny Plaintiffs access to liberal discovery.

Plaintiffs' arguments concerning the reliability of the method are also premature. It is difficult to ascertain that the predictive software is less reliable than the traditional keyword search. Experts were present during the February 8 conference and Judge Peck heard from these experts. The lack of a formal evidentiary hearing at the conference is a minor issue because if the method appears unreliable as the litigation continues and the parties continue to dispute its effectiveness, the Magistrate Judge may then conduct an evidentiary hearing. Judge Peck is in the best position to determine when and if an evidentiary hearing is required and the exercise of his discretion is not contrary to law. Judge Peck has ruled that if the predictive coding software is flawed or if Plaintiffs are not receiving the types of documents that should be produced, the parties are allowed to reconsider their methods and raise their concerns with the Magistrate Judge. The Court understands that the majority of documentary evidence has to be produced by MSLGroup and that Plaintiffs do not have many documents of their own. If the method provided in the protocol does not work or if the sample size is indeed too small to properly apply the technology, the Court will not preclude Plaintiffs from receiving relevant information, but to call the method unreliable at this stage is speculative.³

There simply is no review tool that guarantees perfection. The parties and Judge Peck have acknowledged that there are risks inherent in any method of reviewing electronic documents. Manual review with keyword searches is costly, though appropriate in certain situations. However, even if all parties here were willing to entertain the notion of manually reviewing the documents, such review is prone to human error and marred with inconsistencies from the various attorneys' determination of whether a document is responsive. Judge Peck

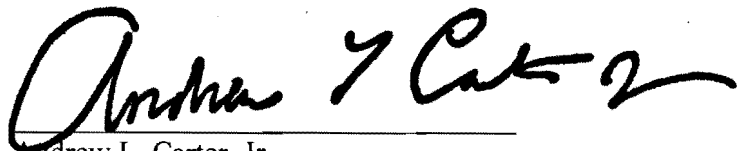
³ The Court adopts Judge Peck's analysis of Rule 26(g) and Fed. R. Evidence 702 for similar reasons provided in his written opinion.

concluded that under the circumstances of this particular case, the use of the predictive coding software as specified in the ESI protocol is more appropriate than keyword searching. The Court does not find a basis to hold that his conclusion is clearly erroneous or contrary to law. Thus, Judge Peck's orders are adopted and Plaintiffs' objections are denied.

The Court also concludes that it should not reject Judge Peck's rulings concerning the production of W-2s. Judge Peck previously ruled that Plaintiffs may visit defense counsel's office and identify the W-2s they want. He also quashed Plaintiffs' third-party subpoena on Automatic Data Processing Services. Plaintiffs and MSLGroup dispute whether the W-2s that Plaintiffs identified have actually been produced. (Compare Pl. Reply at 10 with Def. Opp. at 25.) To the extent that Plaintiffs are missing the W-2s that they identified to MSLGroup, the Court instructs them to send MSLGroup an email noting the missing W-2s within two days of this order. MSLGroup should produce those specified W-2s within three business days of Plaintiffs' email. If there are additional disputes about the missing W-2s, they should return to Judge Peck. Further, Judge Peck's decision to delay production of emails from MSLGroup's chief executive officer until "Phase II" of the ESI protocol is not erroneous or contrary to law. Judge Peck did not make a final determination about whether MSLGroup should produce those emails. The Court reminds the parties that it affords Judge Peck's non-dispositive rulings great deference and that magistrate judges generally have broad latitude with respect to discovery issues.

Dated: New York, New York
April 25, 2012

SO ORDERED.



Andrew L. Carter, Jr.
United States District Judge

VIRGINIA:

IN THE CIRCUIT COURT FOR LOUDOUN COUNTY

GLOBAL AEROSPACE INC., et al.

Plaintiff,

v.

LANDOW AVIATION, L.P. d/b/a

Dulles Jet Center, et al.

Defendants.

CONSOLIDATED CASE NO. CL 61040

CASES AFFECTED

- * Global Aerospace Inc., et al. v. Landow Aviation, L.P. d/b/a Dulles Jet Center, et al. (Case No. CL 61040)
- * BAE Systems Survivability Systems, LLC v. Landow Aviation, L.P., et al. (Case No. CL 61991)
- * La Réunion Aérienne v. Landow Aviation, L.P. d/b/a Dulles Jet Center, et al. (Case No. CL 64475)
- * United States Aviation Underwriters, Inc. v. Landow Aviation, L.P., et al. (Case No. CL 63795)
- * Chartis Aerospace Adjustment Services, Inc. v. Landow Builders Inc., et al. (Case No. CL 63190)
- * Factory Mutual Insurance Company v. Landow Builders Inc., et al. (Case No. CL 63575)
- * The Travelers Indemnity Company, as subrogee of Landow Aviation Limited Partnership v. Bascon, Inc., et al. (Case No. CL 61909)
- * Global Aerospace, Inc. v. J. H. Brandt and Associates, Inc., et al. (Case No. CL 61712)
- * M.I.C. Industries, Inc. v. Landow Aviation, L.P., et al. (Case No. 71633)

ORDER APPROVING THE USE OF PREDICTIVE CODING FOR DISCOVERY

Having heard argument with regard to the Motion of Landow Aviation Limited Partnership, Landow Aviation I, Inc., and Landow & Company Builders, Inc., pursuant to Virginia Rules of Supreme Court 4:1(b) and (c) and 4:15, it 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. with processing to be completed within 60 days and production to follow as soon as practicable and this is without prejudice to a receiving party raising with the court an issue as to completeness or the contents of the production or the ongoing use of predictive coding. in no more than 60 days

23rd

Entered this ___ day of April, 2012.

James H. Chamblin

James H. Chamblin
Loudoun County Circuit Judge

We ask for this:
[Signature]

Jonathan M. Stern (VA Bar No. 41930)
Gordon S. Woodward (VA Bar No. 42449)
Counsel for Lindow Aviation Limited Partnership and
Lindow Aviation I, Inc.

Endorsements waived pursuant to Rule 1:13

Seen and Objected to

W. G. Linton Sr. 4/23/12

William G. Linton Sr
counsel for M.I.C. Industries, Inc.

We ask for this:

Danielle M. LaCoe

Danielle M. LaCoe (USB # 987155)
Attney for Lindow & Company Builders, Inc.

United States District Court
Northern District of California

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

JAYSON HUNTSMAN,
Plaintiff,
v.
SOUTHWEST AIRLINES CO.,
Defendant.

Case No. 19-cv-00083-PJH

DISCOVERY ORDER

Re: Dkt. No. 127

Before the court is a joint discovery dispute letter concerning plaintiff's requests for production of documents ("RFPs") numbers 9-14. Having read the parties' positions and carefully considered their arguments and the relevant legal authority, and good cause appearing, the court hereby **SUSTAINS** defendant's objections.

Background

On February 3, 2021, the court granted plaintiff's motion to certify a national class defined as:

current or former employees of Southwest Airlines Co. who, during their employment with Southwest at any time from October 10, 2004 through the date of judgment in this action, have taken short-term military leave from their employment with Southwest (i.e., military leave that lasted 14 days or fewer) and were subject to a [collective bargaining agreement ("CBA")], except for employees subject to the agreement between Southwest and Transport Workers Union Local 550 covering meteorologists.

Dkt. 95. Now, the parties dispute the appropriate scope of plaintiff's discovery. Plaintiff contends that his RFPs numbers 9-14 seek documents that relate to Southwest's practices for verifying military leave. Defendant disagrees with plaintiff's characterization

1 of these RFPs, posing instead that the requests are both irrelevant and disproportional to
2 the needs of the case. Each of the six requests is considered in turn.

3 **Discussion**

4 **A. Request for Production No. 9**

5 RFP number nine was included in plaintiff's third request for production of
6 documents, served February 12, 2021.¹ Ex. A. Plaintiff contends that this request
7 relates to Southwest's practices for verifying leaves through communications with
8 employees. Defendant argues that this request, seeking evidence unrelated to dates of
9 military service for short-term military leave, overreaches. The court agrees with
10 defendant. Plaintiff is not entitled to discovery of complaints and grievances on all
11 potential USERRA violations given the short-term military leave at issue here.

12 **B. Request for Production No. 10**

13 RFP number 10 was included in plaintiff's third request for production of
14 documents, served February 12, 2021.² Ex. A. Plaintiff contends that this request
15 relates to whether Southwest evaluated Mr. Huntsman's suggestion in 2017 that it
16 conduct a third-party audit of its USERRA policies. Defendant objects, arguing that this
17 request is overbroad and unrelated to the short-term military leave at issue here.
18 Defendant agrees, however, to produce documents that relate to any similar USERRA
19 claim (i.e., an assertion that military leave is comparable to other forms of leave like sick
20 leave, bereavement, or jury duty). The court agrees with defendant's proposed scope of
21

22 ¹ RFP number nine seeks "Documents relating to any complaints or grievances made to
23 Southwest by any Southwest employee regarding Southwest's policies, practices, or
24 protocols for verifying employees' eligibility for leave from work necessitated by reason of
service in the 'uniformed services' (as that term is defined by 38 U.S.C. § 4303(16)),
between October 10, 2004 and the present."

25 ² RFP number 10 seeks "All documents relating to any request or proposal made by
26 representatives of SWAPA and/or members of the Military Resolution Board that
27 Southwest undertake a review of its employment policies for compliance with USERRA in
28 cooperation with an outside party, including but not limited to Southwest's response to
those requests. See Huntsman Tr. 52:21-54:21. This request calls for documents that
would ordinarily fall within attorney-client privilege or the work product doctrine to the
extent that Southwest asserts that the deliberations and opinions of its counsel form a
part of its allegation that Southwest acted in good faith in attempting to comply with
USERRA. See Answer, Fourth & Twenty-Fifth Affirmative Defenses."

1 production—plaintiff is not entitled to discovery on all potential USERRA violations by the
2 airline where the certified class is focused on an alleged failure to pay for short-term
3 leave.

4 **C. Request for Production No. 11**

5 RFP number 11 was included in plaintiff's fourth request for production of
6 documents, served May 7, 2021.³ Ex. B. Plaintiff contends that this request relates to
7 whether Southwest evaluated Mr. Huntsman's suggestion in 2017 that it conduct a third-
8 party audit of its USERRA policies. Defendant argues that this request, in attempting to
9 discover potential knowledge of USERRA requirements more broadly than the short-term
10 pay at issue, is overbroad. However, defendant reports having disclosed seven
11 document custodians whose ESI should be subject to review for responsive documents.
12 Plaintiff argues that the list of seven individuals is incomplete compared to the 13
13 potential custodians he identifies. Defendant's disclosure of a "first-phase" review of ESI
14 is appropriate here, particularly given the large number of responsive documents (over
15 47,000 at this stage). The court agrees with defendant's proposed course of production.

16 **D. Request for Production No. 12**

17 RFP number 12 was included in plaintiff's fourth request for production of
18 documents, served May 7, 2021.⁴ Ex. B. Plaintiff contends that this request relates to
19 Southwest's practices for verifying leaves through communications with employees.
20 Defendant objects to this request as overbroad where it seeks all communications with
21 employees on military leave, regardless of the subject matter or nature of the
22 communication. The court agrees with defendant. Plaintiff is not entitled to all
23

24 ³ RFP number 11 seeks "All electronic communications and attachments thereto that
25 relate to the January 26, 2017 Military Review Board meeting held between Southwest
26 and the Southwest Airlines Pilots Association ("SWAPA") or to the subjects discussed at
27 that meeting."

28 ⁴ RFP number 12 seeks "All electronic communications and attachments thereto that
refer to or reflect presentations, data, survey results, or other information that were used
at or prepared for the November 9, 2016 or January 26, 2017 Military Review Board
meetings held between Southwest and SWAPA, or that were used at or prepared for
Jayson Huntsman's presentation to Southwest personnel in or about June of 2016, which
related to Southwest employees' communications to pilots taking military leave."

1 communications with employees taking military leave without a subject-matter limitation.

2 **E. Request for Production No. 13**

3 RFP number 13 was included in plaintiff's fourth request for production of
4 documents, served May 7, 2021.⁵ Ex. B. Plaintiff contends that this request relates to
5 Southwest's practices for verifying leaves through communications with employees.
6 Defendant objects to this request as overbroad where it seeks all notifications to
7 Southwest that a pilot was taking military leave, regardless of the subject matter or nature
8 of the communication. The court agrees with defendant. Plaintiff is not entitled to all
9 communications with employees taking military leave without a subject-matter limitation.

10 **F. Request for Production No. 14**

11 RFP number 14 was included in plaintiff's fourth request for production of
12 documents, served May 7, 2021.⁶ Ex. B. Plaintiff contends that this request relates to
13 Southwest's practice for verifying leaves through communications with the military.
14 Defendant objects to this request as overbroad and not proportional to the needs of the
15 case, but it still agrees to conduct a phased search of its custodians' data for responsive
16 documents. The court agrees with defendant's phased approach.

17 On this and the other discovery requests, the parties note their disagreement
18 regarding the use of search terms to limit the scope of ESI review. Southwest's approach
19 to using keyword searches and technology-assisted review in tandem does not offend the
20 court's expectation that the parties conduct a reasonable inquiry as required by the rules.

21 _____
22 ⁵ RFP number 13 seeks "All electronic communications and attachments thereto in the
23 possession or custody of the participants in the November 9, 2016 or January 26, 2017
24 Military Review Board meetings held between Southwest and SWAPA or Jayson
25 Huntsman's June 2016 presentation that constitute or relate to communications between
26 Southwest and its employees regarding military leave or to any information exchanged
27 between Southwest and its employees in connection with any request for military leave."

28 ⁶ RFP number 14 seeks "All electronic communications and attachments thereto in the
possession or custody of the participants in the November 9, 2016 or January 26, 2017
Military Review Board meetings held between Southwest and SWAPA or Jayson
Huntsman's June 2016 presentation that constitute or relate to any communications or
any branch of the "uniformed services" (as that term is defined by 38 U.S.C. § 4303(16)),
or with any other government office or agency, including but not limited to the Employer
Support of the Guard and Reserve, regarding Southwest employees' military leaves. The
timeline of this request is April 1, 2016 to present."

Conclusion

For the reasons stated above, the court **SUSTAINS** defendant's objections. The court certified the class on the narrow issue of payment for short-term USERRA leave, not for all potential USERRA violations. The size of the class and the lengthy class period already make for a large and burdensome record of discovery, and the court does not see at this point a need to expand the scope of the case. Plaintiff's requests to compel production are **DENIED**.

IT IS SO ORDERED.

Dated: August 10, 2021

/s/ Phyllis J. Hamilton

PHYLLIS J. HAMILTON
United States District Judge

United States District Court
Northern District of California

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28