IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

JENNIFER LIVINGSTON, et al. Plaintiffs, vs. CITY OF CHICAGO, a municipal corporation Defendant.	Case No. 16-C-10156 Judge Sara L. Ellis Magistrate Judge Young B. Kim

PLAINTIFFS' MOTION FOR COMPLIANCE WITH THE COURT-ORDERED ESI PROTOCOL OR, IN THE ALTERNATIVE, FOR ENTRY OF A PROTOCOL FOR TECHNOLOGY ASSISTED REVIEW

Plaintiffs, by their attorneys, submit this motion for compliance with the current ESI protocol ordered by the Court (ECF No. 239) or, in the alternative, for entry of an appropriate protocol for the use of a Continuous Active Learning ("CAL") computer program to search the email collected by the City. The City should not be permitted to unilaterally apply a CAL program to withhold documents that contain court-ordered search terms without any person reviewing them. Instead, the Court should instruct the City to conduct its responsiveness review using *the City's own ESI protocol* that it filed with the Court when the Court entered the ESI order. ECF No. 219-1 at 6.

Alternatively, if the Court decides to change the existing ESI order and permit the use of CAL, then it should enter a CAL protocol that is technically sound, with necessary safeguards. A CAL protocol should be applied to the correct review population and address initial training sets, quality control, stopping criteria, and validation, among other parameters. To avoid further delay associated with changing course so late in the process, Plaintiffs have attached as Exhibit 1 a technically sound and acceptable CAL protocol that the Court could implement.

Introduction

The Parties began discussing the collection and production of e-mail more than one year ago. From the first meeting, the City proposed, and Plaintiffs agreed, that the City would identify responsive e-mails using keyword search terms and manual review. Never once during the Parties' multiple meet and confers did the City mention the possibility of using technology assisted review ("TAR") or more specifically a CAL program, to search for or identify responsive email. Never once did the City raise this proposal in the briefing and argument of multiple motions, over more than six months, about the ESI parameters, search terms, meetings with IT consultants, or use of a vendor. Now, a full year later, the City is attempting to back out of the Parties' agreements, its own proposed culling and quality control procedures, and this Court's ESI order and to use CAL instead. The City falsely claims that using CAL is simply the "responsiveness review" ordered by the Court. It is not.

It is very late to so radically change the Court's discovery plan, and the Court should not belatedly permit the City to withhold documents based on a computer algorithm without appropriate disclosure and oversight. Even in cases where TAR or CAL are used, the parties create or courts require extensive protocols to ensure the computer program does not fail to identify relevant discovery and that evidence is not concealed. Switching to CAL would require the Parties to implement a new search protocol. The case has already been pending for four years, and Plaintiffs are reluctant to see it delayed further by such a radical departure so late in the process. Nonetheless, in response to the City's preference (indeed insistence) on the use of CAL, Plaintiffs have prepared the attached CAL protocol as an alternative to proceeding as the Court previously ordered.[1] Should the Court decide to deviate from the existing ESI protocol,

[1] Shortly after the last status call, Plaintiffs offered to negotiate a process for the City to use CAL along these terms, but as of filing the City still had not decided whether it was willing

the entry of the attached CAL protocol reflecting accepted technical and other parameters would ensure that the case is not subject to further unnecessary delay.

A. Background.

The long history of ESI disputes in this case is well-documented. The City first approached Plaintiffs about designing an ESI search protocol in March 2019. The Parties negotiated over the issue during several meetings, but by May 2019, the Parties were stuck on the issue of how e-mails should be collected (i.e., by a vendor or using Microsoft Security and Compliance Tool) to ensure that they could be searched effectively. As a result, the Plaintiffs asked the City for a meet and confer between tech representatives. The City refused to participate, so the Plaintiffs filed a motion to compel. ECF No. 157. After the Court granted the Plaintiffs' motion, ECF No. 175, the Parties engaged in the technical meet and confer.

During that meeting, the Parties' tech representatives discussed how to collect e-mails so that they could be searched with keywords. The City's representative acknowledged that its Microsoft Tool would not collect non-OCR'd .pdfs or many Excel spreadsheets that hit on search terms. However, the City nevertheless refused to consider using an outside vendor to conduct an appropriate collection. As a result, the Parties fully briefed the issue of an appropriate e-mail search and review process. ECF Nos. 209, 219, 226.

Among other things, the Parties' briefs addressed the appropriate method to collect e-mail (i.e., outside vendor or the Microsoft Tool) and whether the City should be permitted to review "for responsiveness" the culled e-mails once it applied search terms to the full collection. The City described its proposed responsiveness review as follows:

---

to have those discussions. *See* Exhibit 2 (May 8, 2020 correspondence between Wysong and Schaller).

> First, Defendant would conduct a quality control assessment (typically, by randomly selecting 10% of search results) to confirm the search parameters identify responsive materials. Defendant would then cull the search sample for privileged materials and share the remainder of the sample with Plaintiffs as necessary to refine search terms and parameters. Second, after applying any refinements, Defendant would review the completed search results to cull privileged materials and produce the remaining documents to Plaintiffs.

ECF No. 219 at 14. *See also* ECF No. 219-1 (City's proposed ESI protocol) at 6. The City claimed at that point that the email collection involved more than "4 terabytes" of data, and never once mentioned TAR or CAL in its brief or in the proposed ESI protocol. ECF No. 219 at 2. The City never once mentioned TAR or CAL even though the Seventh Circuit Model Discovery Plan for ESI (on which the City based its protocol), specifically instructs a party to identify whether it intends to use TAR as an alternative to search terms. *Id*. at Ex. 1; Seventh Circuit E-Discovery Pilot Model Order, Sec. B.2.a.(4)(a) ("What search methodology does each Party propose to employ? Search terms? Technology assisted review? Other?"). *See also* N.D. Ill. MIDP Standing Order at C.2.a (directing parties to confer on the "appropriate ESI searches, including custodians and search terms, or other use of technology-assisted review").

In its ruling on the ESI motion, the Court agreed with Plaintiffs that the City's Microsoft Tool was inappropriate for collecting e-mail because it would not permit an appropriate search and required the City to use an outside vendor. ECF No. 239 at 2. It also noted that once the City had collected e-mails and applied initial search terms, the Parties may perform additional *keyword searches* to narrow the initial batch. *Id*. at 5 ("Depending on the number of hits after the initial keyword search using Plaintiffs' proposal, the parties may use more finite terms to reduce the number of hits."). The Court also agreed to allow the City's proposed review process, stating, "[a]s such, the City must be permitted (if it wishes) to review the emails for responsiveness and for privilege before producing them in their native format." *Id*. at 3.

The City has now collected the e-mails and applied the initial court-ordered search terms, resulting in 192,652 unique e-mails. Plaintiffs anticipated that the Parties would then follow the Court's instructions to refine the keyword searches to reduce the number of e-mails so that the City could begin the review process that it described to the Court. However, one day before the Court's April 29 status conference, the City disclosed that, contrary to its previous representations and the Court's ESI order, it did not intend to review the e-mails for "responsiveness." Instead, it intended to review a fraction of the 192,652 e-mails using CAL and then produce some portion of the 192,652 emails that the computer identified and an attorney deemed responsive, without ever looking at the remaining emails. This was the first time that the City ever hinted – much less disclosed – that it planned to unilaterally use TAR or CAL to cull email during the discovery process.

Because CAL is inconsistent with the Court's ESI order, and because it is not appropriate to use CAL after applying keyword searches and without a protocol to provide transparency and safeguards on the operation of the computer, the Parties are at issue.

B. CAL and TAR.

***CAL and TAR Operation*.**

The City seeks to use Relativity Active Learning, a specific Continuous Active Learning or "CAL" tool, which is sometimes called "TAR 2.0." Technology Assisted Review or "TAR" is a general term for using computer algorithms to search large volumes of ESI. In its older initial form, or "TAR 1.0," the parties together use an initial batch or batches of documents marked responsive and non-responsive, the "seed set," to teach the computer algorithm what types of documents are responsive. Once trained, the TAR algorithm searches the full ESI collection and gives each document a classification score from 1 to 100 for the computer's prediction about the

chance the document is responsive. Documents with a high enough classification score are produced (after privilege review). No one looks at documents with lower classification scores.

With CAL, or "TAR 2.0," the initial training set is used to start the computer algorithm on the right track and then the computer presents documents, one by one, to a human reviewer to be marked responsive or not responsive. With each additional document marked, the computer updates the search algorithm with that new information. The computer learns the reviewer's preferences. This is the "active learning" component of CAL. Specifically, the software learns which combinations of text or other features tend to occur in responsive documents and which tend to occur in nonresponsive documents. The software develops a model that it uses to predict and apply responsiveness determinations to unreviewed documents in the overall collection. As a reviewer identifies whether documents are responsive or not responsive, the computer learns what topics are relevant, finds other similar documents in the collection, and pushes them to the top of the pile. The reviewer never needs to review the documents towards the bottom of the pile because the computer has coded the documents as likely not responsive.

By definition, CAL searches for and presents for review only a fraction of documents in a collection. The process continues until the human reviewers hit the "stopping point," or the point of diminishing returns at which continuing with the CAL review is unlikely to uncover many or important additional responsive documents. At that point, if the search is "validated" as sufficiently thorough (as measured by "recall," or the estimated percent of all responsive documents in the collection that have been located), no one looks at the documents not yet presented by the computer.[2]

[2] Put otherwise, TAR in general, is a "process for prioritizing or coding a collection of Electronically Stored Information using a computerized system that harnesses human judgments of subject matter expert(s) on a smaller set of documents and then extrapolates those judgments

Accordingly, and contrary to the City's position, TAR is not used to perform a "responsiveness review." It is an ESI search methodology that performs a function similar to keywords: it narrows down the universe of documents in an e-mail collection so that most documents *never need to be reviewed at all*.

Parties sometimes agree to use CAL on large collections of ESI for two primary benefits. First, with CAL the parties do not need to develop, negotiate, refine, or implement search terms. This saves tremendous time and can, if done correctly with sufficient safeguards, result in a more complete production as well. Traditional search terms might find 50% of responsive documents, while CAL run well can find in excess of 80% of responsive documents.[3] Using CAL, with checks and safeguards, without search terms, may be effective. But using *both* search terms and CAL makes no sense and is not appropriate to meet obligations under the federal rules. If search terms find 50% of the documents, and then CAL is applied to find 85% *of the remaining documents*, then the ultimate production would be just 42.5% of responsive documents (50% x 85%), compared to an 85% recall rate if CAL is used alone.

Second, producing parties want to use CAL because it can save tremendous amounts of time (and cost) because the computer allows the parties to set aside and *never* review large portions of an ESI collection. The risk is that the CAL tool will not be trained properly, or run

---

to the remaining documents in the collection." The Sedona Conference, *The Sedona Conference Glossary: E-Discovery and Digital Information Management*, Fourth Edition, 15 Sedona Conf. J. 305 (2014).

[3] *E.g.* Scott M. Cohen, Elizabeth T. Timkovich, and John J. Rosenthal, Winston & Strawn, *The Tested Effectiveness of Equivio>Relevance in Technology Assisted Review*,Winston & Strawn LLP (Feb. 2014), at 6 (Figure 5), *available at* www.winston.com/images/content/7/8/v4/78010/Winston-White-Paper-eDiscovery-FEB2014.pdf (reporting that an early form of TAR yielded 88.6% of the responsive documents compared to 52.4% for manual review).

long enough, and so large amounts of potentially important ESI could be buried by the computer and never actually reviewed by a human, let alone produced.

CAL Protocols Require Basic Transparency and Safeguards.

The key feature and risk of using CAL (or any TAR tool) is that the computer will use predictive coding to cast aside many responsive documents without review. Courts thus require parties to develop a CAL or TAR protocol that will ensure accuracy, fairness, and transparency. *See Bridgestone Americas, Inc. v. Int'l Bus. Machines Corp.*, No. 3:13-1196, 2014 WL 4923014, at *1 (M.D. Tenn. July 22, 2014) ("Consequently, openness and transparency in what Plaintiff is doing will be of critical importance. . . . The Magistrate Judge expects full openness in this matter."); *Progressive Cas. Ins. Co. v. Delaney*, No. 2:11-CV-00678-LRH-PAL, 2014 WL 3563467, at *4 (D. Nev. July 18, 2014) ("[T]he courts which have allowed predictive coding, and Progressive's own e-discovery consultant . . . have emphasized the need for cooperation and transparency in adopting predictive coding processes and methods."). Without transparency and safeguards from a CAL or TAR protocol, like a traditional ESI search protocol, the process may well veer off track and bury rather than uncover responsive documents. A CAL protocol provides checks and safeguards at key points, including:

a. ***Initial Responsiveness Training***: CAL only works if the reviewers mark documents correctly as responsive or non-responsive, and the CAL tool is set on the right course initially. A CAL protocol thus includes a process to ensure the parties agree on the scope of responsiveness and provide the CAL tool an initial set of documents correctly marked responsive/non-responsive.

b. ***Quality Control***: CAL learns from each mark of responsive/not responsive; errors in this process can lead the computer astray. A CAL protocol thus includes a transparent process to double check the coding decisions during the CAL process.

c. ***Stopping Criteria***: CAL is only effective if run long enough to identify substantially all of the responsive documents in the collection. If it is stopped too soon, many documents are left behind. A CAL protocol thus has metrics by which to identify the "stopping criteria," or likely point of diminishing returns.

d. ***Validation***: Any search must be sufficiently robust and thorough to satisfy Rule 26. This is checked by a process of validation, or calculation of the "recall," that is the percent of responsive documents in the entire collection that have been identified and produced.

At each of these points, there are technical and statistical considerations to the procedures to ensure the CAL process works. With evolving technology risks and benefits, the protocols for running TAR or CAL on an ESI collection can be elaborate, and at times negotiated with court-appointed special masters, for example, in the recent *Broiler Chicken* antitrust litigation where discovery was managed by Magistrate Judge Gilbert. *See* Order Regarding Search Methodology for Electronically Stored Information*, In re Broiler Chicken Antitrust Litigation*, Case No. 16-cv-08637, ECF No. 586 (Jan. 3, 2018) (Exhibit 3).

C. At This Point in the Case, Proceeding Under the Current ESI Order – with the Keyword Refinements and the Responsiveness Review that the City Represented to the Court It Would Conduct – May Reduce Delay.

It has taken a full year for the Parties to negotiate and brief issues for an appropriate ESI search protocol, for the Court to rule on competing proposals, and for the City to collect e-mails and cull them using keywords. Though the Parties have disagreed over specifics of an ESI search protocol, they have *always agreed* that they would identify responsive e-mails through an initial keyword search followed, if needed, by more refined keyword searches. ECF No. 209 at 5-6 (describing Plaintiffs' plan to search email collected by a vendor using search terms); 219 at 3-4 (describing City's plan to self-collect e-mail and search collection using search terms). It is very late now for the City to unilaterally switch methodologies and for a new round of procedure negotiations and dispute resolution. *See Progressive Cas. Ins.*, 2014 WL 3563467, at *11 (rejecting party's late request to use TAR to review e-mails already culled down by keyword search, stating that "approving Progressive's predictive coding proposal . . . will only result in more disputes. It will also further delay completion of discovery in this 2011–filed case."). CAL

is inconsistent with the Court's ESI Order, and if used, it must be done under a technically sound CAL protocol order, with appropriate safeguards.

The City never mentioned CAL or TAR in any pleading filed in the past twelve months, including in its brief delineating procedures for responsiveness review, so it is no surprise that CAL is inconsistent with the Court's ESI Order. ECF No. 239. In that Order, the Court discussed in detail the Parties' respective proposals, always with the understanding that the Parties would identify e-mails through keyword searches. *Id*. at 1 (describing both parties' proposals involving collection and keyword search). If the initial search led to too many results, the City itself proposed pulling a sample of documents for the Parties to review to develop refinements. The Court accordingly noted that if the City's initial keyword search resulted in a universe of e-mails too large to review, then the Parties should narrow that universe by applying more refined keywords. *Id*. at 5 ("Depending on the number of hits after the initial keyword search using Plaintiffs' proposal, the parties may use more finite terms to reduce the number of hits."). The Court never said that the City could search for or narrow the universe of e-mails by unilaterally running CAL on the collection without the Plaintiffs' agreement, without a CAL protocol entered, with no transparency, and without safeguards to ensure that documents are produced.

Nor did the Court approve a "responsiveness review" with CAL. CAL is a way to search an ESI collection to determine what will *not* be reviewed. What the Court approved was the responsiveness review described in the City's brief and set out in its proposed ESI protocol, "a quality control assessment" in which the City reviewed a randomly selected percentage of search results to confirm the search parameters identified responsive materials and turned over the rest of the documents subject to a privilege review. ECF No. 219 at 14; ECF No. 219-1 at 6. That is not a description of CAL or TAR.

Permitting the City to use CAL now – four years into the case – without appropriate procedures and safeguards could easily serve to create more disagreements and result in further delays. Before allowing a computer to set aside large amounts of potentially relevant documents without review, courts require parties to develop a CAL or TAR protocol that will ensure accuracy, fairness, and transparency. *See Progressive Cas. Ins.*, 2014 WL 3563467, at *4 ("[T]he courts which have allowed predictive coding, and Progressive's own e-discovery consultant . . . have emphasized the need for cooperation and transparency in adopting predictive coding processes and methods."); Timothy T. Lau & Emery G. Lee III, Fed. Judicial Ctr., *Technology-Assisted Review for Discovery Requests: A Pocket Guide for Judges* 10 (2017) (attached as Exhibit 4) (hereinafter, "FJC Pocket Guide"). ("[J]udges should urge the parties to agree on a transparent, cooperative TAR process"). In order to avoid further delay, Plaintiffs have attached a technically sound and adequate CAL protocol, which the Court should enter should it decide to replace its existing ESI order with TAR.

Thus, the Court should deny the City's request to unilaterally use CAL with no oversight under the guise of a "responsiveness review." It is inconsistent with the Court's ESI order and will result in delays and invite further disputes at the end of the process. Instead, the Court could instruct the City to follow the responsiveness review procedures set forth in its own ESI protocol (ECF No. 219-1 at 6).

D. Alternatively, the Court Should Enter A CAL Protocol to Search the Entire Email Collection Pursuant to Accepted Procedures With Checks and Transparency.

If the Court permits the City to use CAL at this late date, then it should enter a CAL protocol that is technically sound, with adequate safeguards and transparency, consistent with other courts' approach. *See Bridgestone Americas*, 2014 WL 4923014, at *1 ("The Magistrate Judge expects full openness in this matter."); *Progressive Cas. Ins.*, 2014 WL 3563467, at *4

(requiring transparency in the TAR process); *In re Broiler Chicken Antitrust Litigation,* Ex. 3 (detailing ESI and TAR search procedures). If CAL will be used in this case, Plaintiffs propose a specific protocol with safeguards at each key point in the process: a) initial responsiveness training, b) quality control, c) stopping criteria, and d) validation. *See* Exhibit 1.

There are a few fundamental principles that an adequate CAL protocol should reflect.

First, if the Court approves CAL, then CAL should be applied to the full deduplicated email collection, just as other courts have required.[4] *See FCA US LLC v. Cummins, Inc.*, No. 16-12883, 2017 WL 2806896, at *1 (E.D. Mich. Mar. 28, 2017) ("Applying TAR to the universe of electronic material before any keyword search reduces the universe of electronic material is the preferred method."); *Progressive Cas. Ins.*, 2014 WL 3563467, at *11 (criticizing plaintiff's plan to apply TAR to documents hitting search terms and noting that such a process would be inconsistent with the best practices guide of its vendor); *Rio Tinto PLC v. Vale*, No. 14 Civ. 3042, 2015 WL 4367250, at *1 (S.D.N.Y July 15, 2015) (noting that a party should not cull e-mails using keywords before applying TAR).

Keyword search is notoriously inaccurate when it comes to identifying responsive documents because human beings are not very good at guessing which words will appear in responsive e-mails.[5] The upshot of CAL is that it takes the search term human guessing out of the process. *See Youngevity Int'l, Corp. v. Smith*, No. 16-cv-00704-BTM, 2019 WL 1542300, at

[4] The email collection here could be further narrowed by email threading, which preserves only the last, most-complete message in an email chain, and removes the lesser-included emails. Plaintiffs would also not object to email threading, at the City's option, so long as it does not remove any email with an attachment (since attachments can change when sent back and forth), and the City will produce specific lesser included emails upon request.

[5] *See The Sedona Conference Best Practices Commentary on the Use of Search and Information Retrieval Methods in E-Discovery August 2007 Public Comment Version*, 8 Sedona Conf. J. 189, 202 (2007) ("[keyword searches] are limited by their dependence on matching a specific, sometimes arbitrary choice of language to describe the targeted topic of interest.").

*11 (S.D. Cal. Apr. 9, 2019) ("Predictive coding or TAR has emerged as a far more accurate means of producing responsive ESI in discovery than manual human review or keyword searches."). Applying CAL to the universe of e-mails already subject to search terms would take the worst of both worlds: responsive documents are thrown out with keywords and then missed by the computer. Thus, if the City uses CAL, it should do so on the full deduplicated email collection, as the attached protocol provides.

Second, any CAL process must have essential checks as proposed by Plaintiffs in Exhibit 1. These checks include a process to ensure that the Parties agree at the outset to the responsive/non-responsive coding of the initial set of documents that will train the computer, quality control of coding decisions, and agreement on stopping criteria. *See* FJC Pocket Guide, at9 ("[J]udges should encourage the parties to be transparent in the construction of the seed set."); *Progressive Cas. Ins.*, 2014 WL 3563467, at *10 ("In the handful of cases that have approved technology assisted review of ESI, the courts have required the producing party to provide the requesting party with full disclosure about the technology used, the process, and the methodology, including the documents used to 'train' the computer.").

The City must also be required to validate and disclose how well (or not well) the CAL process found responsive documents. *See Winfield v. City of New York*, No.15-CV-05236, 2017 WL 5664852, at *4 (S.D.N.Y. Nov. 27, 2017) ("Best practices also dictate that the producing party validate the results of its trained TAR system using certain metrics, such as a recall rate that measures the effectiveness of the software in finding responsive documents."); *City of Rockford v. Mallinckrodt ARD Inc.*, 326 F.R.D. 489, 494 (N.D. Ill. 2018) ("Conducting a random sample of the null set is a part of the TAR process [for validation] … Validation and quality assurance are fundamental principles to ESI production. The process provides the reasonable

inquiry supporting the certification under Rule 26(g)."); FJC Pocket Guide, at 11. Validation occurs through the calculation of a recall statistic, which is the number of responsive documents produced, divided by the total number of responsive documents in the collection (as estimated by a sampling procedure). *See, e.g.*, *Broiler Chicken* ESI order (Exhibit 3) (implementing the validation calculation proposed by Plaintiffs).

Plaintiffs' attached protocol would allow City to use CAL following technically sound methodologies, with the basic transparency and safeguards needed to ensure that any CAL process is fair, accurate, and consistent with the City's Rule 26 and 34 obligations.

Conclusion

Wherefore Plaintiffs request that this Court either enforce the current ESI order that requires the City to review emails found by the current search terms using the procedures outlined in the City's proposed ESI protocol or work with Plaintiffs to agree to refinements to the search, or in the alternative, order that any use of CAL shall be applied to the full deduplicated email collection and according to the procedures that Plaintiffs proposed in Exhibit 1 or as otherwise agreed by the Parties.

Dated: May 15, 2020

Respectfully submitted,

/s/ Christopher J. Wilmes
One of the Attorneys for Plaintiffs

CHRISTOPHER WILMES
cwilmes@hsplegal.com
CARYN LEDERER
clederer@hsplegal.com
CHARLES D. WYSONG
cwysong@hsplegal.com
JUSTIN TRESNOWSKI
jtresnowski@hsplegal.com
HUGHES, SOCOL, PIERS, RESNICK & DYM, LTD.

70 W. Madison Street, Suite 4000
Chicago, IL 60602
(312) 580-0100
(312) 580-1994 (Fax)

MARNI WILLENSON
marni@willensonlaw.com
CAITLIN CERVENKA
ckcervenka@willensonlaw.com
WILLENSON LAW, LLC
542 S. Dearborn Street, Suite 610
Chicago, IL 60605
(312) 508-5380
(312) 508-5382 (Fax)

Attorneys for the Plaintiffs