

Exhibit 1

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

**[Proposed] PROTOCOL RELATING TO THE USE OF A
CONTINUOUS ACTIVE LEARNING TOOL (“CAL PROTOCOL”)**

On November 20, 2019, the Court entered an order related to the collection and production of email in this case, covering custodians, time frames, collection procedures, keyword search terms, refinement of keyword searches, review, and production. ECF No. 239. Subsequently, Defendant City of Chicago (“City”) indicated an intent to search a specific collection of email using a Continuous Active Learning (or “CAL”) computer program, specifically Relativity Active Learning (“RAL” or “RAL Tool”). The Court hereby orders as follows to govern the use of RAL by the City.

1. Transparency and Cooperation

To provide checks and transparency without micromanaging how the City meets its discovery obligations, the City will follow these procedures for using RAL stated below. Discovery is a cooperative endeavor and the Parties are directed cooperate and meet and confer as necessary to provide for the production of an adequate number of responsive documents as efficiently as reasonably possible. The Parties will cooperate in good faith to implement this protocol.

2. RAL Review Population

- a. The City will apply the RAL Tool to the collection of email that the City gathered as of May 1, 2020. The collection contains approximately 1.2 million emails, plus attachments (the “RAL Review Population”).
- b. The City will disclose to Plaintiffs the total number of documents in the RAL Review Population and number and nature of documents, if any, i) that it removed from the RAL Review Population for manual review, or ii) that it batch codes without individual review.
- c. The City may remove “junk” emails from the RAL Review Population based upon sender domains disclosed to, and agreed to by, Plaintiffs.
- d. The RAL Review Population may be de-duplicated by hash value across all custodians, with a record (*i.e.*, metadata field) maintained to indicate any custodian(s) who maintained a copy of the document(s) before deduplication. De-duplication may not break apart families. ESI that is not an exact duplicate may not be removed. The City will disclose to Plaintiffs if it elects to use email threading, will not remove any individual email with an attachment, and will produce specific lesser-included emails within the chain upon request.
- e. The City may not apply search terms to the RAL Review Population.

3. Initial RAL Training Set Documents

- a. The City will draw a simple random sample of 1,000 documents from the RAL Review Population (“Initial Training Set”). In addition, either Party may add exemplar or other documents to the Initial Training Set.
- b. The City will review the Initial Training Set and code each document as Responsive or Non-Responsive to Plaintiffs’ requests for production, and Privileged or Non-Privileged; or, if review is not possible for any reason, as Non-Reviewable. For any Initial

Training Set documents coded as privileged, the City will provide Plaintiffs with a privilege log that complies with the Federal Rules of Civil Procedure and also notes whether the document is coded Responsive or Non-Responsive. Plaintiffs will then review and similarly code the non-privileged Initial Training Set documents (“Plaintiff Initial Review Set”). The City will collate the Parties’s codings and create two document logs, one of documents which either Party coded as Non-Reviewable (“non-reviewable log”), and another of documents where the Parties made different responsiveness codings (“conflict log”).

c. The Parties will meet and confer to resolve any conflict in the coding and agree on how to address non-reviewable documents. The City will apply the agreed-upon codings to the Initial Training Set documents to start the RAL Tool. The City will subsequently code documents as responsive or non-responsive consistent with the agreed parameters of responsiveness developed through the coding of the Initial Training Set.

4. Initial RAL Training Process

a. All documents from the Initial Training Set, including those coded as Privileged will be included in the document set used to initially train the RAL Tool’s predictive model.

b. At the start of RAL, the City will report to Plaintiffs the following information about the first 1,000 documents presented by RAL to be reviewed by the City:

- i. the number of responsive documents,
- ii. the number of non-responsive documents, and
- iii. the number of responsive but privileged documents.

c. The City will promptly provide Plaintiffs with a list of these 1,000 documents with their corresponding document IDs and Classification Score, produce the responsive non-privileged documents, and a log of all the documents withheld as privileged.

5. Review and Quality Control

a. The City will code as responsive or non-responsive each document presented by the RAL Tool. The City may not use the Relativity option to “skip” a document. If any document presented for review cannot be reviewed because of a technical difficulty, it should be coded “Technical Difficulty,” presented to Plaintiffs on a technical difficulty log, and not considered part of the RAL Review Population.

b. The City will engage in the following reviewer quality control (“QC”) on a regular, ongoing basis. These measures are used to confirm that the human coding decisions are as accurate and consistent as possible:

- i. The City will take a 10% random sample of documents coded during original review and have a QC reviewer re-review them for responsiveness.
- ii. The City will review all documents not already subject to QC that the RAL Tool gives a “high” Classification Score, but that humans have coded non-responsive. The City will disclose to Plaintiffs the “high” Classification Score it selects for this purpose and the number of documents falling above that score.
- iii. The City will periodically provide Plaintiffs an Overturn Report reporting for each document subject to QC the Classification Score, the original responsiveness call, and the QC reviewer call.

c. The City will manually review each document with a Classification Score higher than the lowest Classification Score of any document marked responsive by any reviewer.

6. Stopping Criteria

- a. The City will review documents identified by the RAL Tool until it meets the following two criteria:
 - i. In each of the 5 most recently reviewed groups of 200 documents, no more than 10% of documents are responsive documents, and none of the responsive documents are novel or more than marginally responsive, and
 - ii. The 1,000 documents in those 5 groups contains no more than 1/6 of the number of responsive documents in the initial 1,000 documents described in Paragraph 4.
- b. The City will report to Plaintiffs the results of the stopping criteria specified herein, including the responsiveness coding and Classification Scores of each of the final 1,000 documents reviewed. The City will also report to Plaintiffs the status of the CAL process including count of documents: (i) in the RAL Review Population, (ii) reviewed to date, (iii) coded responsive and not privileged by humans, (iv) coded responsive but privileged by humans, (v) coded non-responsive by humans, and (vi) coded “Technical Difficulty.”

7. Validation Process

- a. Once the City meets the Stopping Criteria and provides the related reports to Plaintiffs, it will conduct validation according to this paragraph and Appendix A.
- b. The RAL Review Population, less any documents coded as “Technical Difficulty,” will be partitioned into the following three Subcollections:
 - i. Documents identified by the CAL review process as responsive, including any privileged documents, but not including family members of responsive documents, unless those family members are deemed to be responsive in their own right (“Subcollection C(1)”);

- ii. Documents coded as non-responsive by a human reviewer following the application of the RAL Tool (“Subcollection C(2)”); and
 - iii. Documents deemed non-responsive and excluded from manual review as the result of the CAL process (“Subcollection C(3)”).
- c. A simple random sample will be drawn consisting of the following:
 - i. 500 documents from Subcollection C(1) (“Subsample D(1)”);
 - ii. 500 documents from Subcollection C(2) (“Subsample D(2)”); and
 - iii. 2,000 documents from Subcollection C(3) (“Subsample D(3)”).
- d. The sample of 3,000 documents comprised of the documents from Subsamples D(1), D(2), and D(3), will be combined into a single Validation Sample, with no indication of the Subcollection from which the documents were derived or how they were previously coded. The Validation Sample will be reviewed and coded by a subject matter expert (“SME”) who is knowledgeable about the subject matter of the litigation but has not previously reviewed the documents. This should be an attorney who is familiar with the requests for production and the issues in the case. During the course of the review of the Validation Sample, the SME will not be provided with any information concerning the Subcollection or Subsample from which any document was derived or the prior coding of any document. The intent of this requirement is to ensure that the review of the Validation Sample is blind. The SME will code each Validation Sample document as Responsive or Non-Responsive, and as Privileged or Non-Privileged.
- e. The City will provide all non-Privileged Validation Sample documents (“Plaintiff Validation Review Set”) to Plaintiffs. Plaintiffs will review the Plaintiff Validation Review Set, coding each document as Responsive or Non-Responsive. The City will collate the codings of the Parties and create a log of documents where the Parties made different responsiveness codings

(“validation conflict log”). The Parties will meet and confer on documents in the validation conflict log to agree upon the correct responsiveness coding for these documents. The agreed-upon codings will be applied to the Validation Sample documents for the calculation below.

f. Once the agreed coding of the Validation Sample has been completed and reconciled, the City will provide Plaintiffs table listing each of the 3,000 documents in the Validation Sample. For each document, the table will include:

- i. the Bates number or a control/identification number;
- ii. the Subsample from which the document came (*i.e.*, D(1), D(2), or D(3));
- iii. the initial responsiveness coding for the document (*i.e.*, responsive or non-responsive), the SME’s responsiveness coding for privileged documents, and the agreed responsiveness coding for non-privileged documents.

g. Based on the agreed Validation Sample, the City will calculate and report to Plaintiffs the statistics and recall estimate detailed in Appendix A.

h. Once Plaintiffs have received and reviewed the items described in Paragraph 7(f) and Appendix A, the Parties will meet and confer to determine whether or not they agree that the recall estimate, and the quantity and nature of the responsive documents identified through the sampling process, indicate that the review is adequate. If the recall estimate and the samples indicate that Subcollections C(2) and/or C(3) still contain a substantial number of non-marginal, non-duplicative responsive documents as compared to Subcollection C(1), the Parties will meet and confer to discuss potential further action(s) to be taken to identify additional responsive documents either through targeted search, resumption of the CAL process, or some other agreed-upon process. The validation process will be repeated as needed.

i. After validation, the City will produce all non-privileged documents identified as responsive and all non-privileged family members.

8. Additional Considerations

a. To the extent that any aspect of this CAL Protocol turns out to be infeasible or ineffective in practice as applied to a particular category or class of documents, then the Parties are directed to meet and confer to agree on alternative approaches for achieving the stated objective of providing for the production of an adequate number of responsive documents as efficiently as reasonably can be achieved.

b. Should the City determine in its discretion that it does not want to conduct a manual review of any set of documents in the RAL Collection, the City may produce all such documents without review.

c. Any responsive documents that are determined to be privileged will be listed on the City's privilege log unless excluded by agreement of the Parties.

d. The City will produce the non-privileged documents determined to be responsive along with any non-privileged family members in accordance with the ESI production parameters used by the Parties. The production will include an additional metadata field populated with the RAL Tool classification score.

e. In the unlikely event that the Parties cannot agree on the implementation of any step of this protocol, the dispute may be submitted to the Court for resolution.

It is so Ordered:

Date: _____

Hon. Young B. Kim

APPENDIX A

Method of Recall Estimation

An estimate of recall will be computed to inform the decision-making process described in Paragraph 7 of the CAL Protocol; however, the absolute number in its own right will not be dispositive of whether or not a review is substantially complete. Also of concern is the novelty and materiality (or conversely, the duplicative or marginal nature) of any responsive documents identified in Subsamples D(2) and/or D(3). The estimate of recall will be derived as described below. It should be noted that, when conducted by an SME pursuant to Paragraph 7 of the CAL Protocol, a recall estimate over 80% is consistent with, but not the sole indicator of, an adequate review. A recall estimate somewhat lower than this does not necessarily indicate that a review is inadequate, nor does a recall in this range or higher necessarily indicate that a review is adequate; the final determination also will depend on the quantity and nature of the documents that were missed by the review process.

Recall Estimation Method for the CAL Review Process:

The number of responsive documents found \approx the size of Subcollection C(1) \times the number of responsive docs in Subsample D(1) \div 500.

The number of responsive documents coded incorrectly \approx the size of Subcollection C(2) \times the number of responsive documents in Subsample D(2) \div 500.

The number of responsive documents not reviewed \approx size of Subcollection C(3) \times the number of responsive documents in Subsample D(3) \div 2,000.

Estimated recall \approx the number of responsive documents found \div (the number of responsive documents found + the number of responsive documents coded incorrectly + the number of responsive documents not reviewed).