U.S. Department of Justice

Antitrust Division

Liberty Square Building
450 5th Street, N.W.
Washington, DC 20001

This model agreement should not be taken as an indication of any view of the Department of Justice, the Antitrust Division, or any other government department or agency that such practices or procedures are, or should be, legally required. It is not intended to be, and should not be interpreted as, an independent source of substantive or procedural rights or obligations for any party involved in any investigation or litigation with the government or any other individuals or entities. The circumstances of a particular investigation or litigation will dictate whether such an agreement is appropriate; predictive coding is not appropriate in every investigation or litigation. You must consult with Antitrust Division staff before using predictive coding in a particular Second Request investigation.

[Date]

[Attorney]

Re: *[Investigation]*

Dear [Attorney]:

This letter summarizes our conversation on [DATE] relating to your client's methodology for identifying and producing electronic documents that are responsive to the Second Request issued to [COMPANY (or "you" or "your")] regarding the [lNVESTIGATION].

We understand that you plan to use predictive coding software to comply with the Second Request, and your proposed process is attached to this agreement as Exhibit 1. Once collection from all agreed custodians is complete and following deduplication in the manner you described to the Division, the predictive coding algorithm will be run over all collected documents, except as provided below. Based on representations that you made during our calls, and on your agreement to apply the process and validation procedures described below, the Division agrees to your application of predictive coding software.

I. Software Platform and Standards

A. The predictive coding software to be used is [SOFTWARE NAME AND VERSION] from [PROVIDER].

B. The production will meet a recall level of XX% (with an error margin of XX%) and a confidence level of XX%. In addition, a non-responsive sampling analysis will be carried out, requiring a confidence level of XX%, plus or minus X%.

C. Identify the types of metrics available during the training, quality control, and validation process.

II. Seed Set Generation and Training

A. No Analytics Used to Reduce the Review Set. Search terms, manual review, or other analytical tools (e.g., email threading) will not be used to collect documents, or to eliminate documents from the collection prior to deduplication or the application of the predictive coding algorithm.

B. Deduplication should be done vertically (within a custodian's files) and horizontally (across custodians), provided it is done: (1) pursuant to a hash algorithm approved by the Division (e.g., MD5, SHA); (2) prior to application of the predictive coding algorithm; and (3) with the production of a custodian overlay file (updated with each production).

C. Manual Review. Documents that are only found in hard copy, or are uncategorizable (i.e., documents that do not have sufficient text to be categorized using the predictive coding algorithm) will be reviewed manually. If the application of the predictive coding algorithm fails to appropriately categorize these documents, you agree to conduct a manual review of all such documents collected and/or develop an appropriate workflow in consultation with the Division. An agreed list of file types excluded from this process is included in Appendix A.

[OPTIONAL: Specifically, the Division has found that predictive coding, a text-based tool, often is not an appropriate tool to capture responsive materials from files with little text, such as Excel spreadsheets, PowerPoint documents, and images (e.g. PDF) of diagrams or hand-written documents. We understand that you intend to include both Excel and PowerPoint documents in your workflow. If you are unable to verify that your technology can properly categorize these documents, you will conduct a manual review of Excel and Power Point documents.]

D. Subject Matter Experts. Attorneys experienced with all relevant issues arising in the investigation will conduct the review of statistically significant random samples (seed sets) during the assessment and training phases. The following attorneys will train the algorithm (including seed sets, control sets, training rounds and validation rounds): [REVIEWING ATTORNEYS].

E. Foreign Language Documents. Documents that include [XX%] foreign language material must be reviewed using an alternative workflow. They may be reviewed by the predictive coding algorithm, but only if they are part of a separate workflow. That

separate workflow must include distinct seed sets, and native-speaker review of seeds sets and training rounds. Manual review (otherwise consistent with the translation specification) is also acceptable.

[OPTIONAL: You also will exclude from the predictive coding process and review manually documents containing any foreign language content, which will be reviewed by an individual(s) fluent in the language contained in the document.]

III. Work Flow

A. Statistics. You agree to produce to the Division the total number of documents that (1) are collected and ingested, (2) remain after deduplication, (3) are coded responsive, (4) are coded non-responsive, and (5) are uncategorized. Items (3), (4), and (5) will be produced to the Division after each review round.

B. Metadata. In addition to the metadata required by the Division's production specifications, if available, relevance scores for each document produced will be provided.

C. No Supplemental Review for Responsiveness. Except to the extent required to identify privileged information, no manual review or search terms will be used to eliminate documents identified as responsive by the predictive coding process without the written agreement of the Division.

IV. Sample Generation and Selection

A. Sample Generation. Once all training and QC rounds are completed, five (5) random samples will be generated from the pool of non-responsive and non-privileged documents. The Division representative will select one or two samples following the exclusion from those samples of any privileged information.

B. Standards for Validation Samples. The confidence level and confidence interval that will be used to determine the appropriate sample size for a statistically-valid sample must be large enough to account for the removal of privileged documents, and those statistics must be provided to the Division.

C. Duty to Supplement. To the extent that any productions are made prior to the review of the sample by the Division, supplemental productions will be required if there are any changes to responsiveness criteria that are made as a result of the Division's review of the sample.

V. Validation

A. Review. Division representatives will review the sample. This review generally takes less than a day and will be completed in a maximum of three business days from the time that the set is provided for review.

B. Feedback. Division staff will meet with you following the completion of the Division's review of the sample in order to discuss any documents we have identified as responsive and with which of those designations that you may disagree. Depending on the volume and nature of the documents identified by the Division as responsive in the sample, if any, we will discuss at that meeting how and whether any additional review or processing must be done.

Finally, note that the Division continues to improve its procedures to validate the use of predictive coding. This agreement will have no precedential value, and the Division may elect to handle the use of predictive coding in future investigations differently or to apply different standards for an acceptable predictive coding methodology in the future.

Please do not hesitate to contact me if you have any questions.

Sincerely,

Appendix:
Excluded File Types

Exhibit 1:
[The party's final written description(s)]