

EXHIBIT 15

August 20, 2015

VIA ELECTRONIC MAIL

Michael Sobol, Esq.
David Rudolph, Esq.
Melissa Gardner, Esq.
Lieff Cabraser Heimann & Bernstein, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339

Hank Bates, Esq.
Allen Carney, Esq.
David Slade, Esq.
Carney Bates & Pulliam, PLLC
11311 Arcade Drive
Little Rock, AR 72212

Re: *Campbell et al. v. Facebook, Inc.*, N.D. Cal. Case No. 13-cv-05996-PJH

Dear David:

I write in response to your July 23, 2015 letter regarding Facebook's use of predictive coding.

As discussed in our June 19 letter and during our call on July 17, Facebook's predictive coding process is intended to apply advanced machine learning techniques to the text of documents to automatically classify un-reviewed documents into predefined categories of interest, such as responsiveness. The classification models are "trained" through supervised learning—meaning the model is built from a human-reviewed subset of documents—and can be iteratively strengthened with fine-tuning techniques. Our team can then leverage the results to review the documents that are most likely to be responsive so we can more quickly understand and produce responsive content from the document population.

Terminology

As a preliminary matter, we wish to clarify some terminology that is used in your letter but which we believe may otherwise cause some confusion in characterizing how a predictive coding process has been used in this case. Accordingly, for purposes of the explanation below, the term "**training set**" refers to the set of documents that is initially used to "train" the classification model, as described above. The term "**assessment**" refers to a random sample of documents used to evaluate the performance of the classification model—this is often also referred to as a "**QC set**," "**validation set**," or (as in your letter) the "**control set**." The term "**recall**" describes

a measurement used to determine the completeness of the review, and denotes the percentage of all responsive documents that is returned by the model. Finally, the term “**excluded documents**” describes documents that do not meet the requirements for predictive coding and are thus excluded from the process altogether. In this case, Facebook has excluded the following document types from the predictive coding process (in addition, of course, to any file we already reviewed linearly, which may have been used for training or assessment of the predictive coding model as described below). We have no reason to believe that responsive documents are included among these documents.

- .ARC File
- .ZIP/JAR File
- Adobe Indesign Interchange
- Adobe Photoshop
- Compuserve GIF
- Enhanced Windows Metafile
- EPS (TIFF Header)
- EXE / DLL File
- Extensible Markup Language (XML)
- ISO Base Media File
- Java Class File
- JPEG File Interchange
- Macromedia Flash 10
- Macromedia Flash 4-8
- Macromedia Flash 9
- MPEG-1 audio - Layer 3
- MPEG-2 audio - Layer 3
- Portable Network Graphics Format
- Post Script
- Progressive JPEG
- Quicktime Movie
- Tagged Image File Format
- TrueType Font Collection File
- UNIX GZip
- UNIX Tar
- Unknown format
- Windows Bitmap
- Windows Media Audio
- Windows Metafile
- Windows shortcut
- Windows Sound
- Microsoft Digital Video Recording
- MPEG-4 file
- TrueType Font File
- Windows Icon
- Windows Media Video
- Windows Video

Predictive Coding Process

We also wish to provide some clarification as to the nature of the process we have undertaken to conduct predictive coding in this case.

First, we created a training set to teach the computer. The training set includes (i) a set of documents identified as responsive in our linear review, (ii) the results of a review of randomly selected documents from a subset of the overall data set, and (iii) the materials included in Facebook's Production Volumes 3, 4, and 6.¹ As we continue to review and produce responsive documents, those documents will likewise be incorporated into training sets to further train the computer. This method helps to identify enough responsive documents for training where—as here—the responsiveness rate is very low.

Next, we conducted an assessment of the results by reviewing documents in an “assessment set” (the equivalent of what is labelled in your letter as a “control set”). The assessment set is a random, statistically valid, representative sample of the overall data set of unreviewed documents. The statistical parameters used to generate the assessment size were: Confidence level of 95% and Margin of Error of +/- 2.5% and a Variance of 50%.

We then used the training set and the assessment set to generate a predictive model. More specifically, as described during our call, we used a software suite—Equivio—to select documents from the training population until a stable model was created. It then tested the model using the reviewed assessment set. More concretely, in order to evaluate the performance of the predictive model, the assessment was used to evaluate the rate of agreement between the legal team reviewers and the computer.

The predictive model assigned each document in the data set a probability score from 0 to 100, with 0 being the most unlikely to be responsive and 100 being the most likely to be responsive. The coding applied to the documents in the assessment was compared to the probability scores of the documents in the assessment (generated by the predictive model).

“Recall,” as described previously, is a measurement used to evaluate the results of the predictive model, and describes the percentage of all relevant documents that is returned by the model. We added additional documents to the model until it provided a responsive recall of 80%. In other words, the model accurately identified 80% of responsive documents in the assessment set. We intend to review all documents identified by the model as likely to be responsive and produce those that are indeed responsive.

In addition to reviewing the documents identified by the model as likely to be responsive, we will use a “Test the Rest” process to finalize the results of the predictive coding process. “Test the Rest” involves pulling a statistical sample from the documents identified by the model to be unlikely to be responsive and analyzing those documents to confirm that the richness within the

¹ Productions 1 and 2 did not consist of emails, and Production 5 consisted of a single document; these materials therefore were not ideal for inclusion in the training set.

“Rest” is not higher than expected (in other words, the documents with relevance scores below the cut-off do not include an unexpectedly high proportion of relevant documents). The goal is to reconfirm that the responsive recall rate was achieved.

Responses to Additional Statements and Inquiries in your July 23rd Letter

Several statements in your letter concerning our July 17th call do not fully capture our recollection of that call or the relevant facts, and we therefore provide the following clarifications:

- We remind you, as stated during our call, that predictive coding is an iterative process, and as such, if a new population requires review and a linear review cannot be performed efficiently, it may be appropriate to conduct another iteration of the predictive coding process.
- The assessment set (or “control set,” as your letter calls it) actually contains 1,576 documents, not 1,591 documents. This assessment had a point estimate richness of 1.90% (rather than 0.06% as you stated) calculated with a 95% confidence level and 2.5% margin of error.
- Your statement that “Facebook does not intend to further review or produce any of the Filtered By Search Term Documents that fell beneath the cut-off score established during the predictive coding” is incorrect. Facebook will manually review any document that was excluded from the predictive coding process but has a family member with a score above the cut-off.
- Your assertion that “No further training has been performed, although the model produced by the training has been used to classify an unspecified number of additional Filtered By Search Term Documents” also is incorrect. We performed training at multiple stages, including after the most recent upload of documents. Similarly, the assessment set was drawn randomly from all unreviewed “Filtered By Search Term Documents.”

As my colleague Jeana Bisnar Maute mentioned in her email to you on July 27, 2015, we did not agree during our call on July 17th to provide the information you list on page 2 of your letter. Rather, we agreed to engage in a productive discussion and consider your further inquiries with the goal of ensuring that this process is amenable to both parties as a means of making the discovery process efficient and appropriate. We have considered your inquiries and provide the following responses to the questions posed on page 2 of your letter:

1. The total number of documents against which the search terms were run is not readily available using existing tools.

2. We identified the documents to include in the set against which search terms were applied by identifying those documents sent or received by the following former or current Facebook personnel:

Michael Adkins
Jordan Blackthorne
Peng Fan
Dan Fechete
Jonathan Gross
Ray He
Alex Himel
Matt Jones
Mark Kinsey
Ryan Lim
Jiakai Liu
Malorie Lucich
Caryn Marooney
Ben Mathews
Christopher Palow
Giri Rajaram
Scott Renfro
Rob Sherman
Mathew Verghese
Mike Vernal
Frederic Wolens
Gary Wu

3. The number of true positives, true negatives, false positives, and false negatives that resulted from application of the predictive coding model against the assessment set are as follows:

Above the cutoff:

True Positive: 24

False Positive: 654

Below the cutoff:

True Negative: 892

False Negative: 6

We also provide the following additional information in response to your requests:

1. The assessment set is 1,576 documents, comprising a random sample from all unreviewed “Filtered By Search Term Documents” that are appropriate for predictive coding (i.e., not an excluded file type, etc., described above).

2. For information about “whether seeding was used, and if so, when and how,” please see the section above describing the contents and procedure undertaken for the training set.
3. No subset of documents below the cutoff score has been manually reviewed yet (aside from family members of documents above the cutoff score).
4. The former and current Facebook personnel whose documents were included in the roughly 590,000 unique Filtered By Search Terms documents work/ed in the following areas:
 - Social Plugins: Dan Fechete, Jonathan Gross, Ray He, Alex Himel, Mark Kinsey, Scott Renfro, Mike Vernal
 - Communications: Malorie Lucich, Caryn Marooney, Frederic Wolens
 - Messages: Michael Adkins, Ryan Lim, Jiakai Liu
 - Site Integrity: Matt Jones, Ben Mathews, Christopher Palow
 - Advertising: Jordan Blackthorne, Peng Fan, Giri Rajaram, Mathew Verghese, and Gary Wu
 - Policy: Rob Sherman
5. None of the custodians from whom we have collected documents has been excluded from the predictive coding population.
6. The assessment set was classified as follows:

Responsive: 30
Not Responsive: 1,546
7. No documents were unclassified in the training set. All documents were coded. Equivio was able to create a stable model after selecting 1,040 of the documents classified for training. Those documents were classified as follows:

Responsive: 393
Not Responsive: 647
8. The individuals who classified the documents in both the training and assessment sets are Gibson Dunn attorneys.
9. Gibson Dunn attorneys considered the responsiveness of each document and conferred as appropriate.
10. The individuals who classified the documents in both the training and assessment sets are Gibson Dunn attorneys.

11. Gibson Dunn attorneys considered the responsiveness of each document and conferred as appropriate.
12. As described above (*supra* pp. 3-4), both the training and assessment sets contained random samples of documents that were manually classified by Gibson Dunn attorneys, who considered the responsiveness of each document and conferred as appropriate.
13. We have produced such documents, and will continue to produce them, to the extent they are discoverable, responsive, and non-privileged.

As we have explained, Facebook is using these methods to identify the most relevant documents from an enormous set, to expedite the production, and to ensure that there is a fair, reasonable, and proportionate review process. *See* N.D. Cal. ESI Guideline 2.02 (recommending conferring regarding “[o]pportunities to reduce costs and increase efficiency and speed, such as by conferring about the methods and technology used for searching ESI to help identify the relevant information and sampling methods to validate the search for relevant information”).

Please let us know if you have any additional questions.

Sincerely,

/s/ Priyanka Rajagopalan

Priyanka Rajagopalan

cc: All counsel of record (via e-mail only)