

# **EXHIBIT 16**

September 1, 2015

**VIA E-MAIL**

Priyanka Rajagopalan, Esq.  
[PRajagopalan@gibsondunn.com](mailto:PRajagopalan@gibsondunn.com)  
Gibson, Dunn & Crutcher LLP  
1881 Page Mill Road  
Palo Alto, California 94304

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

Dear Priyanka:

Thank you for your letter of August 20, 2015, responding to my letter of July 23, 2015 regarding Facebook's implementation of predictive coding. Based on our understanding of the process Facebook is implementing, we have concerns that Facebook's technology-assisted review ("TAR") may suffer from significant flaws and may not constitute a reliable document search and review process. We are particularly concerned with Facebook's implementation of keyword culling prior to processing, as well as the training process that Facebook has engaged in without input from Plaintiffs.

**Keyword Culling**

As an initial matter Facebook's process of keyword culling—which was not agreed to by Plaintiffs prior to implementation but was instead unilaterally adopted by Facebook—is discouraged and recognized as a flawed methodology that is likely to filter out a significant portion of responsive documents. See *Rio Tinto PLC v. Vale S.A.*, 2015 U.S. Dist. LEXIS 94117 (S.D.N.Y. July 15, 2015) (“[P]re-culling [using keywords] should not occur in a perfect world.”); *Progressive Cas. Ins. Co. v. Delaney*, No. 11-CV-00678, 2014 U.S. Dist. LEXIS 69166, 2014 WL 3563467 (D. Nev. July 18, 2014), (where parties had stipulated to a keyword then manual review protocol, the court would not allow Progressive to use TAR only on the positive keyword hits).

The Equivio process best practices are described in Equivio's “Relevance Project Framework Process Guidelines for Conducting a Predictive Coding Project with the Equivio Relevance Application” (“Guidelines”) (available at [https://law.duke.edu/sites/default/files/images/centers/judicialstudies/Panel\\_1-Background\\_Paper\\_3.pdf](https://law.duke.edu/sites/default/files/images/centers/judicialstudies/Panel_1-Background_Paper_3.pdf)). The Guidelines break down the process into phases for preparation, assessment, training, catch-up (special case of training), decision and verification.

The Guidelines' description of the preparation phase includes the following (p. 2):

2. Metadata culling should be performed prior to loading the data into the predictive coding application . . . In rare cases, basic keyword filters will also be applied. It is recommended that application of keyword filters be minimized as keyword filters are liable to inadvertently cull significant sections of relevant data.
3. Prior to starting the process, it is recommended to engage with opposing counsel, communicating intention to use predictive coding, and verifying the definition of relevance to be applied.

Here, Facebook proposed the use of keyword filters without indicating Facebook's intention to use them in the context of Equivio processing. A primary purpose of TAR is obviate—not simply reinforce the inadequacies of—keyword searching, and courts have recognized that the value of TAR is as an alternative to traditional keyword searches. *See, e.g., Nat'l Day Laborer Org. Network v. United States Immigration & Customs Enforcement Agency*, 877 F. Supp. 2d 87, 110 (S.D.N.Y. 2012) (“[T]he use of keywords without testing and refinement (or more sophisticated techniques) will in fact not be reasonably calculated to uncover all responsive material.”). Keyword searches alone may fail to capture up to 80 percent of relevant documents,<sup>1</sup> and Facebook's proposal of only processing documents returned by keyword searches only exacerbates this problem.

Plaintiffs did not consider, and did not agree to, Facebook's application of keyword filters as a method of culling documents prior to the application of the predicative coding process. Accordingly, in the absence of further information from Facebook regarding the details of the training it has conducted thus far as well as the population of documents against which the keyword searches were run (discussed below), Plaintiffs do not consent to Facebook's current implementation of predictive coding as we understand it.

### **Seeding**

Your letter states “[W]e 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. As we continue to review and produce responsive documents, those documents will likewise be incorporated into training sets to further train the computer.” We interpret this response as indicating Facebook is engaging in “seeding,” which potentially biases the results of the process. As noted in the Guidelines (p. 5):

---

<sup>1</sup> Maura R. Grossman & Terry Sweeney, *What Lawyers Need to Know About Search Tools: The Alternatives to Keyword Searching Include Linguistic and Mathematical Models for Concept Searching*, Nat. L. J. (Aug. 23, 2010).

In most cases, seeding is not used and is not required. Seeding refers to a situation where the user has a set of documents which he knows to be relevant, and which can be fed into Relevance to train the system. Unaided seeding is liable to bias the results; the system will find only documents which are similar to the seed documents, but will not capture other types of relevant documents which may be present, unbeknown to the application, in the population.

Plaintiffs have been given no input into the seeding process; indeed, Plaintiffs are largely in the dark as to the actual composition of the training sets. We request that Facebook: 1) produce and identify by Bates number the “set of documents identified as responsive in our linear review;” 2) produce and identify by Bates number “the results of [the] review of randomly selected documents from a subset of the overall data set” identified in your letter, 3) confirm that all documents produced as Facebook’s Production Volumes 3, 4, and 6 were included in the training set; 4) produce and identify by Bates number any further documents “incorporated into training sets to further train the computer;” and 5) provide the responsiveness classification made by the expert(s) with respect to each seed document.

Without a complete understanding of the documents being used as part of the seeding process, as well as input into that process, Plaintiffs simply cannot judge the effectiveness of Facebook’s TAR implementation at this stage. We also note that there is no provision in the Equivio Guidelines for continuing to introduce new documents manually into the training process post-initial seeding, and request more particulars for the basis of this action, its implementation, and how it fits into the Equivio process.

### **Control Set**

Under the Equivio Guidelines, during the Assessment phase Equivio creates a “control set” of randomly selected documents, which are assessed by the expert(s) and used by Equivio in guiding the training process. Please produce and identify by Bates number those documents, and, with respect to each document in the set, provide the responsiveness classifications made (i) initially by your experts, and (ii) by Equivio after stabilization.

### **Documents Against Which Search Terms Were Run**

In response to our request for “[t]he number of documents against which the search terms were run to produce the initial 600,000 unique documents used to create the predictive coding model,” your letter states, “[t]he total number of documents against which the search terms were run is not readily available using existing tools.” Please explain why Facebook is unable to provide this number. Please also provide 1) a precise listing of the repositories in which the documents against which the search terms were run were stored, and 2) how this document population was selected for keyword searching. Plaintiffs are not in a position to

Priyanka Rajagopalan, Esq.  
September 1, 2015  
Page 4

assess the overall effectiveness of Facebook's TAR implementation or its appropriateness in this instance without this information.

Please provide the requested documents and information by September 4, 2015. Given the impending summary judgment and class certification deadlines, Facebook's production thus far—a significant portion of which are either publicly-available or highly duplicative (i.e. individual responses to email chains spread across many documents)—appears to be inadequate, which may in part be due to Facebook's failure to implement best practices in its predictive coding process. For example, Facebook has yet to produce any of the documents related to Facebook's decision to scan private messages for URLs, and to increase the "Like" count for third-party websites as discussed in Hank Bates' August 20, 2015 letter.

If you have any questions about or would like to discuss the foregoing, please let us know.

Sincerely,

A handwritten signature in black ink, appearing to read 'DTR', with a long, sweeping flourish extending to the right.

David T. Rudolph

DTR/wp  
1271660.1