Embracing E-Discovery in Antitrust Matters: Slow But Steady Progress Toward Convergence between the U.S. and the UK?

Lawyers are sometimes risk adverse and slow to change. Although this tends to lead to a more cautious approach to embracing new technologies, including the use of artificial intelligence, the increasing burden of e-discovery has forced the issue. Lawyers on both sides of matters increasingly are embracing the rise of a technology known as “predictive coding” to identify responsive and nonresponsive documents in private litigation and government investigations. While the United States is on the leading edge of this trend, other jurisdictions, including the United Kingdom, have been slower to follow suit, particularly in antitrust matters.

This is a timely and important issue. Recent research shows that nearly half of the cases requiring UK electronic corporate data to be processed were either in preparation for, or in response to, UK or foreign antitrust and regulatory matters. This dynamic has led to predictions that lawyers in the UK (and elsewhere) are expected to make greater use of artificial intelligence in the near future.

The U.S. experience is illustrative. The two federal antitrust agencies—the U.S. Department of Justice (“DOJ”) and Federal Trade Commission (“FTC”)—have agreed with parties that predictive coding is useful to cull large volumes of electronically stored information in antitrust investigations. By contrast, there has not been any clear statement on this subject from the UK’s Competition and Markets Authority (“CMA”), the UK sector regulators, or the courts. That changed in February 2016, when the High Court of England and Wales for the first time endorsed the use of predictive coding in the UK, relying in large part on judicial acceptance of the technology in the U.S. (Case No: HC-2014-000038, Pyrrho Investments Limited and another v MWB Property Limited, [2016] EWHC 256 (Ch)).

This Commentary discusses the latest trends in the use of predictive coding in U.S. and UK antitrust matters, and how Pyrrho is likely to spur slow but steady progress toward greater acceptance in the UK.

What Is Predictive Coding?

Use of e-discovery tools to alleviate the burdens associated with document-intensive matters is not new. Since the mid-1980s, private litigants have agreed to use keyword searches, “concept-based” searches, and most recently predictive coding as alternatives to manual document-by-document (“linear”) review. Generally speaking, predictive coding is a form of artificial
intelligence that uses human reviewers’ examination of a subset of documents (so-called “seed documents”) to “train” computer algorithms to review and “predict” what other documents are responsive. Nowadays, the term “predictive coding” is used interchangeably with “technology-assisted review” (“TAR”), “computer-assisted review,” or simply “assisted review.”

How Does It Work in Practice?

There are a number of different software platforms capable of performing the necessary analytics for predictive coding. Lawyers work with e-discovery vendors to understand the capabilities of the predictive coding software to ensure that the document population is handled appropriately. For example, some predictive coding models cannot categorize certain file types, which would need to undergo a linear review. However, other predictive coding software platforms do not have the same limitations. In terms of training protocols, there are two broad categories of how predictive coding models can be trained. In “passive learning” protocols, the model is trained by evaluating multiple sets of random samples of documents coded by attorney reviewers. In “active learning” protocols, the computer helps select certain “borderline” documents for attorneys to review to further refine the model more efficiently than entirely random document sets would.

While the various software platforms may employ assorted processes and have varied limitations, a key objective across all of them is the ultimate “recall rate”—i.e., the percentage of relevant documents ultimately discovered—that is validated by the nonresponsive sample results. An agreed-upon recall rate allows litigants, merging parties, and government agencies to vet the effectiveness of the predictive coding platform regardless of the software used.

Use of Predictive Coding in U.S. Antitrust Merger Investigations

In the U.S., the use of predictive coding is becoming standard practice in response to the significant compulsory document requests (“Second Requests”) issued by the federal antitrust agencies to parties in antitrust merger investigations. Increasingly, law firms are engaging with the DOJ and FTC on behalf of their clients to use predictive coding to identify responsive documents. Employing this type of technology is becoming necessary to handle the growing volume of emails and other electronically stored information that companies generate and to comply with often stringent time limits dictated either by the merger review process or the deal timetable (or both).

Take, for example, the DOJ and FTC processes. The DOJ has amended its “Model Second Request” to require merging parties to disclose and discuss any “software or technology used to identify or eliminate potentially responsive documents and information produced in response to this Request, including … predictive coding.” If a merging party chooses to use predictive coding, the DOJ and the party typically will agree to a certain recall rate and the opportunity for the DOJ to review statistically significant samples of (nonprivileged) nonresponsive documents to verify the agreed-upon recall rate. The DOJ has generally accepted a 75 percent recall rate with at least a 90 percent confidence level, which acknowledges that no review will be perfect and that approximately 25 percent of the responsive documents will not be produced.

The FTC’s position is largely the same. Citing the widespread use of electronic materials and the need to improve the efficiency of its investigations when proposing changes to its rules of procedures, the FTC has stated, “Document discovery today is markedly different than it was only a decade ago…. Searches, identification, and collection all require special skills and, if done properly, may utilize one or more search tools such as advanced key word searches, Boolean connectors, Bayesian logic, concept searches, predictive coding, and other advanced analytics.” Accordingly, in August 2015, the FTC also amended its Second Request Model to include instructions on the use of predictive coding.

This embrace of new technology by the DOJ and FTC is encouraging, as a growing body of evidence has demonstrated that linear reviews, in which attorneys review each document one-by-one for responsiveness, are less accurate and generate recall rates well below 75 percent.

Although the U.S. antitrust agencies are leading the way in accepting the use of predictive coding in antitrust matters, the agencies’ electronic discovery negotiations are not internally consistent, and the path can sometimes be challenging, depending on the agency staff assigned to the matter. If not
reasonably managed (on both sides), discussions about the e-discovery process can last weeks, elevating process over substance, delaying forward progress on the merits of the investigation. Some agency staff may attempt to evaluate how the software is performing before the model is fully trained and to set detailed parameters for the process that are not always or obviously consistent with best practices. As one of the leading U.S. judicial voices in e-discovery, Magistrate Judge Peck, cautioned, “one point must be stressed—it is inappropriate to hold TAR to a higher standard than keywords or manual review. Doing so discourages parties from using TAR for fear of spending more in motion practice than the savings from TAR for review.”

The U.S. merger review process puts the burden on the merging parties to certify that they “substantially complied” with a Second Request, and, in such a context, the merging parties should retain broad discretion to use the method they reasonably believe to be appropriate, proportionate, and effective in order to satisfy their duty to comply. The agency staff can evaluate the sufficiency of the documentary response by verifying the agreed-upon recall rate through the review of samples of nonresponsive documents. As Judge Peck explained, “requesting parties can insure that training and review was done appropriately by other means, such as statistical estimation of recall at the conclusion of the review as well as by whether there are gaps in the production, and quality control review of samples from the documents categorized as nonresponsive.”

**UK Comparison**

In the UK, predictive coding has been used in litigation and antitrust matters, but not as often as in the U.S. For example, in a high-profile alleged price-fixing conspiracy investigation some years ago, the CMA’s predecessor—the Office of Fair Trading—agreed to the Jones Day antitrust team’s request to use predictive coding to identify a few thousand responsive documents from an original collection dataset of several million documents, with a return rate and confidence level similar to those accepted by the DOJ in “Second Request” merger investigations. Despite some experience with the technology, there continues to be no formal statement or guidance from the CMA or other UK sector regulators. This can perhaps be attributed to two reasons: UK antitrust matters, particularly merger reviews, have tended to be less document-intensive than U.S.-style “Second Requests,” and there has been a historic reluctance by the English courts (and consequently, lawyers and government agencies) to endorse the use of predictive coding. But things are changing.

First, since April 2014, the CMA’s investigatory powers have been strengthened, in particular through a wider power to require parties to produce documents at all stages of an investigation, including during first-phase merger reviews and market studies, and the ability to impose significant financial penalties on parties that fail to comply with an information request notice without a reasonable excuse. UK sector regulators have similar powers.

Second, and crucially, in February 2016, the High Court of England and Wales officially approved the use of predictive coding for the first time in the UK in *Pyrrho*.

**Pyrrho Ruling**

This case concerned compensation claims from the shareholders of a company on various grounds, including breach of fiduciary duties. The court ordered the disclosure of all relevant documents. Initially, the total number of electronic files was “more than 17.6 million.” After de-duplication, the total was narrowed to 3.1 million documents, which the court observed was “still a large and costly number to search.” The parties turned to predictive coding to expedite the review and asked the court to approve this approach. The court noted that “there is not a great deal by way of guidance, and nothing by way of authority, on the use of such software as part of the disclosure process.” Such a lack of authority prompted the court to analyze other jurisdictions and to draw comparisons with the well-known U.S. district court decision of Da Silva Moore in which Judge Peck endorsed predictive coding for the first time in judicial proceedings (see our previous *Antitrust Alert* on the topic).

In approving the use of predictive coding in *Pyrrho*, Master Matthews listed the following 10 reasons in favor of the use of predictive coding and found “no factors of any weight pointing in the opposite direction”:
Other jurisdictions have found that predictive coding software is useful in appropriate cases, notably the U.S. (Da Silva Moore).

There is no evidence that predictive coding is less accurate than linear review (and indeed, there is evidence that it is more accurate).

There is greater consistency in using predictive coding over "dozens, perhaps hundreds, of lower-grade fee-earners, each seeking independently to apply the relevant criteria."

There are no prohibitions on the use of predictive coding in the applicable rules of procedure.

The number of electronic documents to be reviewed in this case was "huge, over 3 million."

The cost of manually searching these documents would be "enormous, amounting to several million pounds at least." The court even goes further to describe a manual review of each document as "unreasonable" where a "suitable automated alternative exists at lower cost."

The costs of using predictive coding would be a fraction of the cost of manual review.

The value of the claims made in the litigation are in the tens of millions, making the estimated cost of predictive coding proportionate.

If the predictive coding is unsatisfactory, there will still be time to consider alternative methods.

The parties have agreed on the use of the software and a protocol.

In his closing remarks, Master Matthews noted that the agreed protocol was case specific: "Whether it would be right for approval to be given in other cases will, of course, depend upon the particular circumstances obtaining in them."

Implications

Pyrrho was not an antitrust case, but it is nonetheless instructive for the use of predictive coding in UK antitrust matters in at least two respects:

First, Pyrrho has been hailed as a victory for proportionality: The use of predictive coding may be appropriate in circumstances where it is effective in ensuring that disclosure exercises remain proportionate. This is particularly important for UK antitrust matters, where the CMA and the other UK sector regulators are under a duty to make sure that each request for information is justified and proportionate and enables companies to balance their duty to cooperate with the exercise of their rights of defense. The principle that document disclosure requests must be proportionate has recently been reaffirmed in Cases C-247-268/14 P. Italmobiliare and Others—in which the Court of Justice of the EU found that the European Commission's requests for information directed at cement manufacturers in an EU antitrust probe were "extremely numerous" and excessive, and thus annulled such requests. Although these cases relate to European Commission investigations under EU competition law, they are also relevant in principle for the application of UK antitrust rules.

Second, in Pyrrho, the court was satisfied that training and review were done appropriately without the need for the disclosure of the "seed" set of documents to the other side. This is particularly relevant for antitrust investigations. Like U.S. Second Requests, it is for the parties that are subject to a disclosure request from the CMA or other UK sector regulators to certify compliance with such a disclosure request. Therefore, there is a strong argument that parties should remain free to use other reasonable means for vetting the accuracy of their disclosure, such as the statistical estimation of recall at the conclusion of the review based on the quality control review of samples from the documents categorized as nonresponsive and nonprivileged.

Toward Convergence between the U.S. and the UK

Lawyers in the UK are likely to rely on the Pyrrho judgment in the future in support of the use of predictive coding in response to a large document disclosure requests, including in antitrust investigations.

Given that e-discovery is one area where the U.S. is leading the way (as recognized by Pyrrho itself), some guidance could usefully be drawn from DOJ and FTC experience, with a view to achieving a consistent approach to the use of predictive coding in U.S. and UK antitrust matters. Accordingly, in deciding whether it may be appropriate to propose the use of predictive coding in an antitrust investigation, the parties and their lawyers should take into account the following considerations:
Volume, Timing, and Collection Logistics. Consider whether predictive coding is the most efficient solution after evaluation of the document volume and collection logistics. Predictive coding may not save time and money if the volume of documents is low or if documents have to be collected and processed in small, incremental batches.

Experience. Consider whether the investigating agency staff has experience with predictive coding. The CMA and some of the UK sector regulators are increasingly using predictive coding for the prioritization and review of documents disclosed to them in response to information requests. A less-experienced agency staff may be less likely to agree to predictive coding or, alternatively, more inclined to challenge or delay accepting a certification of completeness where predictive coding has been used without advance acceptance by agency staff. In the absence of formal guidance or additional precedent in antitrust investigations in the UK, the conditions set out in Pyrrho for accepting predictive coding provide a useful precedent.

Recall Rate. Make sure that you are comfortable with the recall rates. Even if the agency staff has agreed to the use of predictive coding, you will still be required to certify the efficacy of the methodology and substantial compliance with the document disclosure request.

Methodology and Protocol. Consider what aspects of the methodology and protocol will require prior agreement with the agency. A highly transparent protocol could complicate the review and open the door for an expanded and time-consuming inquiry, especially if the agency staff does not have a good understanding of the technology or visibility of what information is contained in the documents at the outset. The goal of any review process is to return a satisfactory volume of responsive documents, and ultimately the burden rests on the party deploying the technology to use it appropriately to reach the desired recall rate—which the agency can validate through nonresponsive samples. Agreement on the recall rate and verification of that rate through the review of nonresponsive samples should instead be sufficient to endorse any review process without the unnecessary distraction of prolonged discussions regarding the specific software and work flows.

The Continued Need for Some Linear Review. Linear reviews of predicted responsive documents that contain potentially privileged communications, as identified by “privilege” search terms, are still common as parties seek to identify 100 percent of privileged communications. But as technology and legal standards advance, parties to an investigation and their lawyers should keep an open mind and be prepared for further change and development in this area.

Lawyer Contacts
For further information, please contact your principal Firm representative or one of the lawyers listed below. General email messages may be sent using our “Contact Us” form, which can be found at www.jonesday.com/contactus/.

Ryan C. Thomas
Washington
+1.202.879.3807
rchomans@jonesday.com

Francesco Liberatore
London
+44.20.7039.5221
fliberatore@jonesday.com

Ausra O. Deluard
San Francisco
+1.415.875.5852
adeluard@jonesday.com