

Reproduced with permission from Digital Discovery & e-Evidence, 16 DDEE 283, 06/23/2016. Copyright © 2016 by The Bureau of National Affairs, Inc. (800-372-1033) <http://www.bna.com>

### TECHNOLOGY ASSISTED REVIEW

Kroll Ontrack's Michele C.S. Lange and Tracey Stretton write to broaden readers' understanding of all-things global when it comes to predictive coding, to help legal teams—in any country—prioritize and categorize documents for review with greater efficiency, more accuracy and less cost.

## Predictive Coding: Taking the Globe by Storm



By MICHELE C. S. LANGE AND TRACEY STRETTON

**B**eing two steps ahead is elemental to a successful legal practice. From unearthing every fact and relevant case law opinion to anticipating the opponent's next argument, legal teams thrive on keeping unexpected storms at bay.

*Michele C.S. Lange is the director of thought leadership and industry relations for Kroll Ontrack. In this role, she regularly writes and speaks on topics related to electronic discovery, computer forensics and technology's role in the law. [mlange@krollontrack.com](mailto:mlange@krollontrack.com)*

*Tracy Stretton is a legal consultant for Kroll Ontrack. She advises lawyers and their clients on the use of technology in legal practice and has provided consultative expertise on a large number of cases in a variety of international jurisdictions. [tstretton@krollontrack.com](mailto:tstretton@krollontrack.com)*

*Anna Guler, Kroll Ontrack law clerk, assisted with the development of this article.*

Nobody can afford to be caught off-guard, and this is equally true when it comes to technology. Across the globe, legal professionals need to understand how technology can improve their ability to represent their clients, while at the same time staying abreast of how big data is influencing their client's actions, behaviors and legal obligations.

For legal professionals looking to stay vigilant, predictive coding is one such advancement in technology that is providing a port in a storm of big data sets.

Attorneys, paralegals, legal IT professionals, and litigation support staff need to understand what predictive coding is and more importantly, how it is evolving around the globe. The “must have” information includes:

- How are parties using predictive coding technology in the U.S., Europe and Asia?
- What global nuances must legal professionals understand when predictive coding technology crosses continents?
- What are courts saying about predictive coding in America and how does that differ from judges' opinions in the U.K. and other countries?

### Predictive Coding is Transcending Borders

No matter the language or location of data, predictive coding—also known as technology assisted review (TAR) or computer assisted review (CAR)—helps cull through mountains of data by leveraging algorithms to quickly identify relevant and responsive documents, making the document review process less time consuming for legal teams.

**Importance of Training.** But, predictive coding does not work in a vacuum. Human attorneys must interact with and train the system by reviewing a sample of the overall document set to empower the machine to learn how to rank and sort documents. It is this human interaction with a powerful algorithm that ultimately allows technology assisted review to work in any country, in any language and, in any type of matter.

In fact, the quality of the entire predictive coding process is only as good as the training it receives. Without adequate training by subject matter expert attorneys, the technology will not learn how to properly make categorization assessments or suggestions.

**Size Matters.** In the early planning phases of a document review, it is central to assess whether the case will benefit from the predictive coding process.

Experience shows that predictive coding is best applicable in cases with large data volumes—hundreds of thousands of documents.

While case size is a valid factor, it is worth bearing in mind that legal teams have successfully applied predictive coding to document sets with as few as 5,000 documents.

Other important factors when considering predictive coding include:

- the complexity and languages of the documents in the case;
- the availability of subject matter experts to train the system; and
- 

time pressures and deadlines.

As a party in any matter or in any country is looking to apply predictive coding, these same factors must be considered.

**Need for Expertise.** Lastly, having a trusted expert that is experienced in the predictive coding process is also a best practice that applies no matter the geographic location of the case.

Until a party has deep experience with the technology, legal teams should seek the assistance of a predictive coding consultant, expert employed by a law firm, or technology vendor to provide guidance along the way.

This becomes even more imperative when a predictive coding project involves parties or evidence located globally.

## A Global Predictive Coding Case Law Primer

As predictive coding technology is embraced by legal professionals, attorneys turn to the bench for insight. Opinions from judges regarding cases that use predictive coding technology are critical to the development of best practices.

More importantly, attorneys must keep their eye on case law developments around the world as many of their clients likely operate on a global scale.

In February 2012, Magistrate Judge Peck of the Southern District of New York, issued the first opinion written in the United States on the use of computer assisted review technology. *Da Silva Moore v. Publicis Groupe* holds the stamp of approval for the use of pre-

dictive coding technology in the litigation process.<sup>1</sup> Magistrate Judge Peck states:

The decision to allow computer assisted review in this case was relatively easy — the parties agreed to its use (although disagreed about how best to implement such review). The Court recognizes that computer assisted review is not a magic, Staples-Easy-Button, solution appropriate for all cases. The technology exists and should be used where appropriate. . .<sup>2</sup>

More recently in March 2015, Magistrate Judge Peck issued another opinion regarding the use of computer assisted review. In *Rio Tinto PLC v. Vale*, the court once again approved the parties' use of computer assisted review.

However, unlike in *Moore v. Publicis Groupe*, the parties in *Rio Tinto v. Vale* mutually agreed upon the predictive coding protocol to be used before bringing the issue to the court.<sup>3</sup>

However, these are not the only two cases that address the use of predictive coding technology within the United States. There are over a dozen other cases within the U.S. that issue formal written opinions regarding predictive coding, in which the courts evaluate the key issues of predictive coding—whether its use is appropriate, how it should be used, and whether disclosure is necessary.

**View From Across the Pond.** American jurists are not the only ones evaluating the applicability and use of predictive coding technology in discovery.

In March 2015, Ireland followed the lead of the U.S. with its approval of the use of predictive coding in the discovery process, aptly named “disclosure” in the United Kingdom, with “ediscovery” the corollary term to “ediscovery.”

In *Irish Bank Resolution Corporation Ltd. & Ors v. Quinn & Ors*, the Irish Bank Resolution Corporation (IBRC) calculated the length of time and cost that would be associated with the review of thousands of documents. Realizing the amount of resources it would take, the IBRC proposed the use of predictive coding and created a protocol to address the defendant's concerns.

However, the defendant did not consent to the use of such technology.

The Irish High Court held that in the discovery of large data sets, technology assisted review using predictive coding is at least as accurate as, and probably more accurate than, the manual or linear method of identifying relevant documents and allowed the plaintiff to use the TAR protocol they created.<sup>4</sup>

English courts are not far behind the U.S. and Ireland in their approval of computer assisted review for litigation. In February 2016, Master Matthews issued the first British opinion that approved the use of predictive coding in High Court proceedings, partly relying on Judge Peck's opinion in *Moore v. Publicis Groupe*.

In *Pyrrho Investments Ltd. v. MWB Property Ltd.*, the bulk of relevant documents recovered from backup tapes, which contained data from e-mail accounts used by the second to the fifth defendants, were in the con-

<sup>1</sup> *Da Silva Moore v. Publicis Groupe*, 287 F.R.D. 182 (S.D.N.Y. 2012)

<sup>2</sup> *Id.*

<sup>3</sup> *Rio Tinto PLC v. Vale S.A.*, 306 F.R.D. 125 (S.D.N.Y. 2015)

<sup>4</sup> *Irish Bank Resolution Corporation Ltd & Ors v. Quinn & Ors* [2015] IEHC 175

trol of the second claimant. The overall number of documents recovered totaled more than 10 million.

One of the defendants proposed that the claimant use predictive coding to cull through the documents as the ediscovery process was left for the parties to decide.

The parties agreed on the use of predictive coding and Master Matthews approved, reasoning that there is nothing in the Civil Procedure Rules or Practice Directions prohibiting the use of predictive coding and that such technology has been proven to be less costly and more accurate (as evidenced by cases in the U.S. and Ireland).<sup>5</sup>

In May 2016, Britain had its first *contested* case regarding predictive coding, *David Brown v. BCA Trading*.

David Brown wished to adapt a “traditional” document review approach, while BCA Trading, who held the majority of the documents, sought to apply the approach taken in *Pyrrho Investments v. MWB* arguing that a traditional approach would be excessive and not proportionate to costs.

David Brown’s counsel argued that predictive coding should not be used as such technology is untrustworthy and would lead to a number of relevant documents being missed.

BCA’s counsel on the other hand countered that the use of predictive coding was more reliable and trustworthy.

The High Court ultimately ruled in favor of BCA Trading and allowed the use of TAR, relying on the statistic that traditional review would cost two and a half times more than employing the use of predictive coding.

---

**APAC, EMEA, and Canada are jurisdictions to watch for future predictive coding decisions from courts, as this technology is increasingly adopted.**

---

**Tip of the Iceberg?** These opinions are just the start of what we will see on the issue of predictive coding in years to come. Case law will continue to develop as many of the same concerns are being raised globally.

Due to the similarity of the issues being raised around the world, borders are blurring when it comes to case law opinions in respect to predictive coding. Courts are looking to judges in other countries to help guide their decisions, regardless of the differing case types and applicable laws.

Case law opinions in the United States have led to more standardized predictive coding practices between parties, and as more cases arise in the United Kingdom, they are likely going to follow the U.S. in creating guidance on which parties can rely.

Furthermore, APAC, EMEA, and Canada are jurisdictions to watch for future predictive coding decisions from courts, as this technology is increasingly adopted in those regions.

---

<sup>5</sup> *Pyrrho Investments Ltd. v. MWB Property Ltd.* [2016] EWHC 256

## Evolving Global Perceptions, Misconceptions

As predictive coding takes root globally, legal teams have very similar concerns and misperceptions about the technology and how it works in document review.

**Is It Effective?** For example, effectiveness is a significant concern facing predictive coding teams no matter the geography in which it is being used. In the U.S. opinion *Moore v. Publicis Groupe*, Magistrate Judge Peck acknowledged the imperfection that comes with computer assisted document review, but emphasized that perfection is not the goal and that the focus should be on using a process that is reasonable and proportional to the matter.<sup>6</sup>

Despite the extensive U.S. judicial support of technology assisted review, legal teams in America are still somewhat hesitant to leverage predictive coding due to widespread belief that human review is the gold standard.

This misperception is not unique to American document review teams. However, various studies prove the belief to be untrue.<sup>7</sup>

The bottom line is that humans are inconsistent (both with one another and with themselves). This reality is magnified when you have a large number of document reviewers working on a single document set.

When predictive coding is used to categorize documents, the machine does not get tired or distracted, and when it learns something new, it goes back and re-reviews documents to make better decisions with new information.

**Disclosures of Protocol, Training Docs.** Another common roadblock to the uptake of predictive coding technology relates to whether or not a party must seek approval from its adversary or disclose its predictive coding methods, including the protocol and documents used to train the system.

Again, this concern crosses borders and is not only an issue for U.S. legal teams applying predictive coding. Parties are unsure of how transparent they must be with respect to the predictive coding processes they are applying and there are few bright line guideposts.

If parties cannot come to an agreement, there is nothing currently in the U.S. or U.K. case law that would require a party using predictive coding to turn over all of the information they used in training to the requesting party, without prior agreement to do so.

Specifically, in the U.S., prevailing wisdom is that parties control how they search for and find relevant documents in discovery; as long as a party conducts a diligent search, its review methods are considered work product.<sup>8</sup>

If the requesting party is concerned about the producing party’s predictive coding processes, Judge Peck stated that there are other measures that the requesting party can take to ensure accuracy in training and review, such as statistical estimation of recall at the conclusion of the review or quality control.<sup>9</sup>

In British courts, there are some guidelines that are set out in the *eDisclosure Protocol*. Litigants in British

---

<sup>6</sup> *Da Silva Moore*, 287 F.R.D. 182, 191

<sup>7</sup> *TREC Legal Track*, Text Retrieval Conference (May 23, 2012, 11:11 PM), <http://trec-legal.umiacs.umd.edu/#about>.

<sup>8</sup> *Rio Tinto*, 306 F.R.D. 129

<sup>9</sup> *Id.*

courts must keep detailed records of the process applied to their documentation from identification and collection onwards in order to provide a trail for what process has been applied to each document category, including the methodology and logic used to remove any documents from collection.<sup>10</sup> However, Master Matthews states:

I add that the judges of the Technology and Construction Court support an eDisclosure Protocol, produced by practitioners and available on the website of the Technology and Construction Solicitors' Association. This does contemplate the use of predictive coding software in appropriate cases. But it is only a protocol and has no normative force.<sup>11</sup>

In the absence of many guidelines, Master Matthews applies a similar approach to that used by Magistrate Judge Peck and the Irish courts. He leaves the predictive coding protocol for the parties to discuss and agree upon and does not make any statement that would require the parties to turn over information about their training processes.

**Language Barriers?** Lastly, as predictive coding is being deployed on a global scale, legal teams do face a few deeper-level concerns that most parties need not worry about.

The most worrisome concern is language detection. In the U.S., a common misconception is that predictive coding technology can only be used to detect the English language and if used to detect information in other languages, it will not work or it will be inaccurate. However, nothing is further from reality.

Technology is language agnostic, and predictive coding can be trained for use with any language as long as the legal team retains trainers fluent in the particular languages in the document set. This may take some trial and error, but it is possible to use the system with multiple languages and have it be just as accurate as when it is used solely with the English language.

## Predicting Coding In Action

To illustrate the pervasiveness of predictive coding technology around the world, consider this case study: A U.S.-based manufacturing company operated with its key suppliers and assembly units based in Japan. Management and sales were based in America, resulting in many documents—such as manufacturing specifications, invoices, and bills of lading—to be multilingual and frequently transmitted via e-mail between the U.S. and Japan.

When a products liability suit ensued in a U.S. court, this global company was obligated to produce English

<sup>10</sup> *TeCSA / SCL / TECBAR eDisclosure Protocol*, The Technology and Construction Solicitors' Association (Jan. 9, 2015), <http://www.tecsa.org.uk/sites/default/files/eDisclosure%20Protocol%20Version%2002%20-%2009%20January%202015.pdf>.

<sup>11</sup> *Pyrrho Investments Ltd. v. MWB Property Ltd.* [2016] EWHC 256

and Japanese documents in discovery. The company's legal team desired to leverage predictive coding to speed the document review, but because of the multilingual document set, concerns abounded.

**Training Issues.** For instance, one issue that the legal team encountered was training. The predictive coding algorithm had to be trained to sufficient quality levels in both English and Japanese in order for the system to prioritize documents and suggest categorizations.

The legal team worked closely with its eDiscovery service provider to source a small group of document reviewers fluent in Japanese. These attorneys sat alongside the English speaking attorneys and both teams served as trainers for the predictive coding engine.

Effectiveness metrics had to be examined for both the English and the Japanese training sets before the predictive coding engine's recommendations could be confidently leveraged.

**Syntax Challenges.** In addition to language fluency, legal teams in global predictive coding matters can run across multilingual syntax challenges.

Predictive coding algorithms are text-based engines. They are not able to detect dialects or contextual differences across languages.

For example, between various languages, common words or phrases have different meanings or are used in a different contexts.

Different regional spellings is another common challenge when training predictive coding engines with multilingual data sets. As an example, 'color' in American English is spelled differently than 'colour' in British English. Since predictive coding is text based (identical to any typical search), documents with multilingual spellings or contextual differences should be found and trained within the predictive coding process.

The main takeaway is that when using TAR for multilingual document reviews, the training team should be attorneys who are not only fluent in the languages involved, but also attorneys who understand different dialects, phrase meanings and common spelling mistakes inherent in the document set.

## What's Next on the Horizon?

In the U.S. and U.K., judges will continue to interpret the nuances of parties' predictive coding practices.

In other European and Asian countries it is only a matter of time before a court in those regions formally addresses predictive coding, meaning predictive coding will become as ubiquitous as traditional linear search tools.

In the meantime, technology is evolving all the time and as data volumes and types grow, so too does the sophistication of predictive coding tools.

To stay ahead of competitors and meet the expectations of demanding international clients working across languages and jurisdictions, legal professionals must remain vigilant amidst the predictive coding storm in looking for the next technological developments.