Man or machine?

Faras Baloch charts the potential role of artificial intelligence in disclosure & privilege review in criminal cases

IN BRIEF

- How computers 'learn' from humans.
- The potential complications of large datasets.
- What we can expect from AI in the future.

Artificial intelligence or 'AI' generally describes a computer solving problems in a way that a human might. It encompasses 'machine learning' which involves a computer classifying data and learning when it makes mistakes.

Machine learning algorithms have a vast number of applications which include providing personalised recommendations in search engines, identifying unwanted emails through spam filters, and assisting medical practitioners diagnose illness.

Increasingly, lawyers are harnessing this technology to assist in the review of large datasets. In this context, the term 'machine learning' is used interchangeably with 'Technology Assisted Review' (TAR) and 'Computer Assisted Review' (CAR).

How does it work?

It works in a similar way to an email spam filter. The spam filter categorises emails as either 'spam' or 'not spam' and removes the spam emails to a separate folder. Every time the user flags an email as spam or not spam, the computer learns from that decision. It determines whether the presence of a feature makes it more or less likely to be spam. For example, it may learn that the presence of the word 'bitcoin' in an email increases the probability of it being spam by 80%, but the word 'presentation' decreases the probability by 50%. Over time, millions of features are assigned probabilities and the spam filter predicts with accuracy whether an email is spam or not.

A basic spam filter uses a 'Naïve Bayes' classification algorithm or 'classifier' which is commonly used in computer assisted reviews.

Problems created by large datasets

The proliferation of digital devices and expansion of data storage capabilities has had a huge impact on criminal investigations. Prosecutors can face enormous challenges when having to review digital material.

A serious fraud investigation can typically involve a prosecution team having to examine a 10TB dataset—the equivalent of 65 million pages or 13,000 filing cabinets full of paper. Even if search terms and other tools successfully exclude 90% of the likely irrelevant material, this still leaves 6.5 million pages to review.

Usually, a team of disclosure counsel is engaged to determine relevance and disclosure. A team of ten disclosure counsel who never fall ill, take holidays, or make mistakes might review 6.5 million pages in around five years. They will almost certainly spend most of the five years reviewing irrelevant material because search terms generate a vast number of false positives.

Claims of privilege can add further complications. To ensure that privileged material is not released to the prosecution, a team of specialist independent counsel must be retained to review material responsive to privilege search terms. Again, they will spend most of their time looking at non-privileged material due to the many false positives generated by search terms.

How can the classifier help?

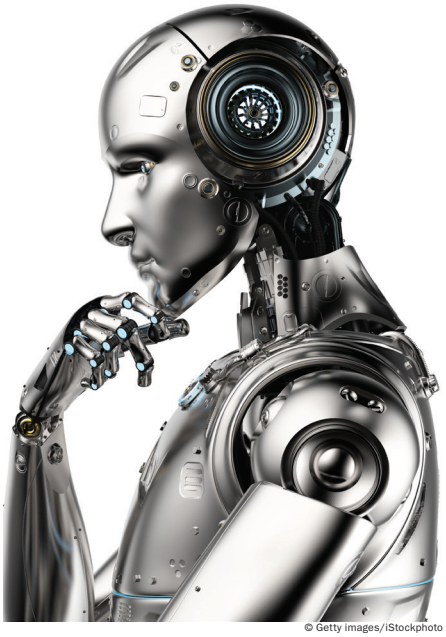
The counsel teams are, essentially, categorising material. Disclosure reviewers categorise material as 'relevant' or 'not relevant'. If it is 'relevant' they go on to consider if it is 'disclosable' or 'not disclosable'. Privilege reviewers categorise material as 'privileged' or 'not privileged'.

Just as the spam filter learns from users flagging an email as spam, the classifier learns from counsel's decisions. It breaks down the features of each item counsel has reviewed and in respect of each one calculates whether its presence in a document increases or decreases the probability of it falling into a particular category. Counsel need only review a small proportion of the overall material, perhaps around 2%, which serves as 'training data' for the classifier.

Once trained the classifier is deployed to review part of the dataset—to put into practice what it has 'learnt'. Its categorisations are checked by counsel and where it has made mistakes these are used to retrain and refine it. This step is repeated until the classifier reaches an acceptable level of accuracy. It can then review and categorise all the material with little human input.

The classifier reveals how confident it is in each decision. The features of a particular document may point to two mutually exclusive categories, in which case, it may be less confident about the categorisation. That document can be reviewed by counsel and the decision can be fed back to provide more 'training data'.

The benefit of a classifier over manual review is that it saves time and cost. A review of 6.5 million pages may take around five months and not five years. Unlike a human review the cost does not increase at the same rate as the number of pages to be reviewed



© Getty images/iStockphoto

increases. Doubling the page count doesn't double the time or cost.

Studies in the US have shown that a classifier yields more accurate results than manual review. Judge Peck of the Southern District Court of New York noted in one judgment that 'while some lawyers still consider manual review the 'gold standard', that is a myth'.

What problems may be encountered?

The accuracy of the classifier depends upon the quality of the training data it is supplied by human reviewers. If the quality of decision making is poor the classifier will make poor decisions. In other words, 'garbage in, garbage out'. If human reviewers are inconsistent in their decision making the classifier is likely to fail. That could happen if there are too many reviewers, they are insufficiently experienced or have not been given clear instructions.

The reliability of the classifier also depends on the quality of the software being used. Earlier versions were entirely reliant on training data and were unable to adapt through trial and error. As a result, they were less reliable.

Unusual datasets such as those containing numerous foreign language items, handwritten documents or non-standard file types can also cause delays in the classifier acquiring the desired level of accuracy.

Has it been used before in criminal cases?

No contested criminal case has used a classifier for disclosure or privilege review. The Serious Fraud Office's (SFO's) investigation into Rolls Royce, which concluded in 2017 with a deferred prosecution agreement, made use of a classifier to identify privileged material among around 30 million documents.

In 2018 the SFO announced that it would use 'Axcelerate' an AI powered document review system in all new casework. However, it is unclear whether the SFO intends to deploy it in disclosure or privilege reviews.

Attorney General's Guidelines

The Attorney General's Disclosure Review in 2018 recognised that the growth of digital material was 'outpacing human capacity to handle it'. It suggested that a 'different approach, researching and developing appropriate solutions using predictive coding or Artificial Intelligence was needed in high end cases'. The review supported the use of these solutions with further detail to be set out. The Attorney General's Guidelines on Disclosure came into force on 31 December 2020. At para 41 they provide that: 'Technology that takes the form of search tools which use unambiguous calculations to perform problem-solving operations, such as algorithms or predictive coding, are an acceptable method of examining and reviewing material for disclosure purposes'. While approval is given to the use of algorithms such as the Naïve Bayes classifier, no further guidance has been provided.

Judicial consideration

No court in England has given a judgment which has considered the use of AI for disclosure or privilege review in a criminal case. However, the use of this technology has been considered in relation to disclosure in civil cases and some useful principles can be derived.

In *Pyrrho Investments Ltd v MWB Property Ltd & Ors* [2016] EWHC 256 (Ch) the parties agreed to its use, but sought the court's approval. Master Matthews approved of the use of a protocol which defined the dataset, sample size, confidence level and margin of error for the classifier. He emphasised the requirement for accuracy and consistency in the training data observing 'there will be greater consistency in using the computer to apply the approach of a senior lawyer towards the initial sample (as refined) to the whole document set, than in using dozens, perhaps hundreds, of lower-grade fee earners, each seeking to independently apply the relevant criteria in relation to individual documents'.

In *Triumph Controls UK Ltd and another company v Primus International Holding Co and other companies* [2018] EWHC 176 (TCC) the claimants did not provide relevant detail about how the technology was set up or operated or how the sampling exercise was conducted. Ten paralegals and four associates were used to create the training data, but there was no overseeing lawyer (as recommended in *Pyrrho*). Mr Justice Coulson described the exercise as neither transparent nor independently verifiable.

Courts in the US, Ireland and Australia have embraced the use of classifiers in large scale document review exercises. In all jurisdictions the courts have: (i) required the parties to agree a comprehensive protocol for the use of the technology; (ii) emphasised the importance of transparency between the parties; (iii) recommended the use of senior lawyers to create the 'training data'; and (iv) highlighted the role of regular quality assurance reviews to test the classifier's decisions.

The future

Storage capacity doubles roughly every four years so the challenges faced by law enforcement agencies will become more difficult to overcome unless technology is harnessed effectively. Artificial intelligence is part of the solution, but it is no silver bullet. It won't be suitable in every case. Where it is used, humans will still play a crucial role in reviewing data, quality assurance and agreeing protocols. However, a future in which small teams of experienced lawyers spend a few months creating training data and quality assuring a classifier is, in my view, more likely and more desirable than ever-expanding teams of disclosure counsel spending years in a manual review. NLJ

Faras Baloch is a barrister at Red Lion Chambers (*www.redlionchambers.co.uk*).