Digital Justice and the Use of Algorithms to Predict Litigation Outcomes[1]

INTRODUCTION

The use of artificial intelligence ("AI") is now a familiar feature of large-scale commercial litigation and arbitration, insofar as predictive coding or technology assisted review[2] is accepted as a means of complying with the obligation to conduct a reasonable search for documents responsive to disclosure orders. This development has been a steady one since its application in English proceedings was foreshadowed by Senior Master Whitaker[3]. Nevertheless, the first judgment approving the use of predictive coding in English proceedings was given only two and a half years ago[4]. There has so far been no mainstream use of the technology beyond the disclosure context.

Predictive coding is essentially a labour and cost saving tool which automates the narrow task but time consuming process of searching for relevant documents in electronically stored information (typically the largest part of any modern disclosure process). This paper considers the prospects of a different, and potentially more far-reaching, use of artificial intelligence in law: predicting outcomes of cases. It considers the state of the existing technology, the current (limited but illuminating) research, and what the future might hold from the perspective of litigants (and their funders and insurers), lawyers, and the wider justice system. Those future possibilities are also used as a way to think about some of the challenges the technology might face in a legal context.

---

[1] This paper is a revised version of a talk given by Charles Ciumei QC at the Bar European Circuit Annual Conference in Stockholm on 21 September 2018. The author would like to express his thanks to Benedict Tompkins for his help in its preparation.

[2] Also called 'computer assisted review'.

[3] *Goodale v Ministry of Justice* [2009] EWHC B41 (QB)

[4] *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch), see in particular ¶17-24 for a description of predicting coding. In the USA, the technology seems well-established and the Irish High Court approved the use of predictive coding in *Irish Bank Resolution Corporation Ltd v Quinn* [2015] IEHC 175.

THE TECHNOLOGY

The kind of AI discussed in this papers is so-called ‘weak’ AI, in the sense that it is focussed exclusively on a very specific task and is in contrast to the still somewhat distant perfection of general AI that can apply itself to any given problem. We are surrounded by weak AI without even being aware of it: For example in our smart phones. The facial recognition that unlocks my phone is powered by a neural network running on an specially optimised processor. The software readily recognises my face, but if I position my glasses on the bridge of my nose and peer over them in the fashion of a sceptical judge, suddenly I am a stranger to it and my phone won’t unlock. That is ‘weak’ AI, albeit it is also immensely powerful and ‘works like magic’ at least some of the time.

Before diving in, it may be helpful to start with some definitions. In this context ‘machine learning’ is a subset of AI and means a computer program that has the ability to improve its performance at given task (i.e. learn from its experience) without being explicitly programmed (i.e. without changes to the program being made by a human). For the purposes of our discussion, there are broadly two types of techniques: (i) statistical approaches (such as decision trees and support vector machines); and, (ii) neural networks.

Although the technology is complex and its features may not be comprehensible to legal or lay users, lawyers will need to have at least a conceptual understanding of the tools (as any solicitor who has drafted a witness statement dealing with predictive coding will know). A key point to understand is that during its training and deployment the machine learning model is not analysing the law or applying rule-based reasoning familiar to lawyers, which has the effect that the results produced by the model may very well not be open to inspection or interrogation in the traditional way.

A very simplified description of the basic process which is used to prepare and deploy a machine learning tool to predict the outcome of litigation is as follows:

1. **Gather and prepare data to be used to train the machine learning model.** The data used will depend on the task envisaged. In predictive coding of documents for disclosure purposes, it is a sample of the universe of documents, reviewed and tagged for relevance (etc) by a human reviewer. For

predicting the outcome of litigation, to date the academic studies have used as raw material judgments of previous cases. This includes, as well as the outcome, features such as the facts of the dispute, the identity of the judge, the identity of the parties, the lawyers involved, etc. The dataset is then described by "features", which are identified as having predictive value. These features may not be comprehensible to lay audiences, but in legal contexts will invariably be text-based. For example, in the ECHR study by *Aletras et al.*[5], considered further below, the features used from the sample judgments were "N-grams" (sequences of linguistic units of defined lengths, derived from a text treated as a "bag-of-words", or the text stripped of grammar, syntax and word order – effectively a proxy for semantic content) and "topics" (statistical clusters of related N-grams).

2. **Select a model to train.** There are a large number of already coded computer models that can be used for classifying things and making decisions about them. So this is really a matter of choosing an existing model and adapting it for the purpose at hand. The academic studies considered below used statistical models; commercial users (such as litigation funders) may not disclose the nature of the model employed.
3. **Train the model.** The model is then trained by iterative exposure to the dataset: it considers a judgment and predicts the outcome. To date the research has determined outcome on a binary basis: win/ no win; violation/ no violation. If it gets the answer wrong, parameters in the model are changed until more answers become correct. Crucially, the training process is automatic and the adjustment of parameters is done by the model itself (hence, "machine learning"). Training stops when a pre-determined level of accuracy is reached.
4. **Use the model.** Once the outcome is good enough for the particular task, it can be deployed to predict the decision of 'new' data (for example, decided cases that were not in the training data, or undecided cases). The model then

[5] Aletras et al (2016), Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. **PeerJ Comput. Sci. 2:e93; DOI 10.7717/peerj-cs.93**

makes predictions about the outcomes in new cases that it has not previously seen before.

ACADEMIC RESEARCH: DOES IT WORK?

Academic research into how judges make decisions is a long-standing field of enquiry. However, in the last two decades or so, researchers have applied machine learning tools to their enquiry. Three studies are considered which show apparently remarkable powers of prediction compared to the highly trained legal mind.

In the first study, *Ruger et al.*[6], the authors took as their training data all decided cases of the Rehnquist US Supreme Court prior to the 2002 Court term, and described those decisions by reference to six variables selected as "*plausibly correlated with outcomes*" (liberal or conservative direction of the lower court decision, issue area in the case, circuit of origin, type of petitioner, type of respondent, and whether it was argued that a practice is unconstitutional). The study then used a model consisting of classification trees (a statistical technique) to consider cases pending before the Court in the 2002 term, presented to the model on the information available prior to oral argument. By way of a comparison, a group of expert lawyers were asked to predict outcomes for the same set of 2002 cases. The result was a clear win for the machine: the model predicted the correct outcome with 75% accuracy, whereas the expert lawyers achieved 59.1%.

In their 2016 study, *Aletras et al.*[7], did not pit human against machine. Nor did they use their model to predict the outcome of undecided cases. Rather they applied the model to large number of previously decided cases, but which decisions were 'new' to the model as they were not part of the training data set. They trained a machine learning algorithm using a data set of 594 European Court of Human Rights decisions on Articles 3, 6 and 8 of the Convention (described using N-grams and topics, as explained above), equally balanced between

---

[6] Ruger T.W., Kim P.T., Martin A.D, & Quinn K.M., The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking *Columbia Law Review* **104**, 1150-1210 (2004)

[7] See above.

decisions finding a violation of the Convention and those finding no violation. They used a support vector machine, which is also a statistical approach. Following training and testing, the model could predict the correct outcome (violation or no violation) based on the text of judgments 79% of the time.

Building on the earlier work of Ruger et al., in 2017 the basic features of the earlier US Supreme Court study was updated, but with a far larger dataset and machine learning technology. *Katz, et al.*[8] used as its raw material a database of US Supreme Court decisions between 1816 and 2015, consisting of 28,009 cases and 243,882 votes of individual judges. That dataset was then described using a highly complex set of features. The model used was a decision "forest", which once trained, predicted the overall case outcome with 70.2% accuracy, and individual justice votes at 71.9% accuracy.

Although not an academic study, the methods employed in this academic research have also recently been demonstrated in a commercial setting. Late last year CaseCrunch, a UK startup, held a competition between its predictive algorithms and various teams of London commercial lawyers, to predict the outcome of a sample PPI mis-selling cases decided by the Financial Ombudsman Service[9]. The algorithm got it right 86.6% of the time, and the lawyers 62.3% of the time.

APPLICATIONS AND LIMITATIONS

Those results, which are already impressive, are likely to improve. The issues turned to now are the potential applications, and the technical limitations and principled problems that may be encountered in such applications.

The commercial appeal of accurate outcome prediction is obvious. Insurers want to know which cases they are going to win and lose, both in order to make settlement decisions and to effectively allocate capital. Litigation funders need to be able reliably to pick winning cases in order to make a return on their investment. Lawyers require an appreciation of the likely outcome to advise

[8] Katz DM, Bommarito MJ, II, Blackman J (2017) A general approach for predicting the behavior of the Supreme Court of the United States. PLoS ONE 12(4): e0174698. https://doi.org/10.1371/journal.pone.0174698

[9] www.artificiallawyer.com/2017/10/28/ai-beats-human-lawyers-in-casecrunch-prediction-showdown/

clients and to make strategic decisions. And both funders and lawyers could deploy machine-based predictions in negotiations not only in an absolute fashion (by being confident of a particular outcome), but in a relative fashion (by noting the superiority of a technology-assisted view of the merits over the traditional variety). There is some anecdotal evidence of the former already. There are also litigation consultants already established in the USA (and at least one in the UK) that offer to provide strategic litigation insight based on the use of AI techniques.

The Ministry of Justice, as well as public and judicial bodies in general, are likely to be interested in the potential of any such technology. Although the prospect of a robot judges may seem like a remote one, it is worth bearing in mind that substantially automated dispute resolution processes already exist: For example, in the consumer sphere, on eBay. In any event, a more achievable aim might be the development of prediction tools to assist human judicial decision making. The authors of the ECHR study, for example, suggest that their model "*can ... be used to develop prior indicators for diagnosing potential violations of specific Articles in lodged applications and eventually prioritise the decision process on cases where violation seems very likely*". In other words, decision-making at the admissibility stage of ECHR proceedings (at which the vast majority — about 84,000 in 2015 — of cases are sifted out) could be assisted by artificial intelligence, both saving time and therefore cost, and improving the quality of the system by speeding up final determinations. In the current climate, anything that offers the prospect of reducing the costs of the justice system is likely to attract the attention of policy makers. There is some jurisprudential precedent for such a formulaic approach to judicial decision-making. In *United States v Carroll Towing*[10] Judge Learned Hand famously reduced liability in negligence to an equation: "*if the probability [of loss] be called P; the [gravity of the resulting] injury, L; and the burden [adequate precautions], B; liability depends upon whether B is less than L multiplied by P: i. e., whether B > PL*". Thus expressed, certain categories of case at least appear amenable both to automated decisions and highly accurate automated predictions.

[10] *United States v. Carroll Towing Co* 159 F 2d 169 (2d Cir., 1947). That said the suggestion that the application of the burden of proof might be approached as a mathematical exercise was recently disapproved by the Court of Appeal: *A (Children)* [2018] EWCA Civ 1718, 25 July 2018.

There is also the potential for sizeable effects at the most basic level of the economics of litigation. Economic theory suggests that cases go to trial only when the parties' respective views of the merits mean that there is not a settlement range acceptable to both parties. And in turn, the parties' views of the merits depend on the information (including, crucially, legal advice) available to them. As one leading text puts it: "*a mutually beneficial settlement exists as long as the plaintiff's estimate of the expected judgment does not exceed the defendant's estimate by more than the sum of their costs of trial*"[11]. By increasing the accuracy of outcome prediction (and, possibly, reducing information asymmetry between parties with access to different calibres of legal advice), algorithmic legal tools can be expected to increase settlement, which may be good for legal consumers, if not lawyers. Of course, this may depend to an extent on the validity of the assumption in economic theory that litigants are rational actors, whereas for understandable reasons on the part of the litigants, this is not always the case.

However, a future of machine-aided (or machine-driven) lawyering and judgment presents problems, both technical and philosophical:

1. Machine learning models are **highly constrained and task-specific**. This is why the first and now most established legal application is predictive coding of disclosable documents. The research summarised above is similarly constrained to single courts of carefully circumscribed judgment databases. The content of the judgments is highly structured and a world away from the much less structured material deployed at the trial or hearing. It will be less of a stretch to apply the techniques to simple or high-volume cases (financial mis-selling, or road traffic accidents, for example), but more complicated commercial disputes will be more difficult. Models will also very likely need adjusting when moving between courts and jurisdictions. And as the subject matter moves further away from the core of relatively straightforward applications, the challenges will increase: commercial arbitration, for example, lacks the comprehensive datasets needed to train models; and pre-

[11] Steven Shavell *Foundations of Economic Analysis of Law* (Harvard University Press, 2004) at p.403

dispute setps (such as settling pleadings or evidence) lack the binary win–lose outcomes that are the other side of the basic process summarised above.

2. **Bias** is an inherent risk of the technology if not intelligent thinking in general. Machine learning algorithms must start with something, and if the training data is biased the outcome will be affected. The most high profile concern in this area is the possibility of machine-assisted discrimination (for example, if part of the sentencing process were to become automated), but it also gives rise to accuracy problems: if the training data is mostly composed of cases that lose, the algorithm may predict more losing cases because it has seen insufficient winning cases.
3. **Transparency and reasons**. Machine learning algorithms are, for the most part, not open to interrogation. They are to a large extent 'black boxes'. This means that while the models may produce accurate answers, they are not necessarily capable of explaining the route from the inputs to those answers: and certainly not in a way that is particularly meaningful or accessible. They are not rule-based reasoning systems. This is not a concern where the only goal is prediction (litigation funders do not care *why* a case is a winning case, just that it is likely to win; a disclosing party does not care *why* a document is relevant and therefore disclosable, just that it is). But it may become an acute problem if the technology is used not only to predict but also to decide. In particular, a 'black box' outcome is not amenable to appeal in the sense of examining the defensibility of an outcome against external criteria (the law and the evidence). The implication is that appeals against such automated decisions might have to be as of right and effectively re-hearings, which might undermine any costs saving.
4. **Outliers.** Current technology is good at what it does (and increasingly so), most of the time. But these systems in general do not deal well with outliers. In a different field, but one with the most severe consequences in the event of error, some current self-driving vehicle systems may sometimes have difficulty reacting to stationery objects ahead when travelling at speed, in part

because such events are infrequent[12]. Fortunately, in applications where getting the right answer most of the time is an acceptable result, this may not be a debilitating flaw — again, disclosure and litigation funding are two such applications. However, our concept of justice involves, at least in theory, getting the right result even in highly unusual cases, and those cases are often important to the development of the law. The outlier cases are often the cases that cause a change in the direction of the law. In relation to predicting case outcomes, the outlier problem is related to the bias problem: in most circumstances, the training data will come from decided cases, but cases that reach court (or, in the ECHR example, the very small minority of cases that are admitted) are the exception rather than the rule. This is already an established problem in law and economics: as Shavell notes (at p.433) "*the cases that go to trial may be very different from the population of cases that settle, so that generalizing from trial cases is difficult and may be misleading*". Widespread use of machine learning technology has the potential greatly to amplify that issue.

5. **Societal acceptance.** The deployment of such systems beyond the use of private parties and in order to automate or assist judicial decision making brings wider concerns into play. Our children may be content to have their legal disputes adjudicated by chatbot, but are there wider implications? Society as a whole benefits from the orderly and peaceful adjustment of disputes. That extends beyond mere judicial decision making. Recently in R v Unison[13], the UK Supreme Court delivered a strong rebuke to the notion that justice can be valued in solely monetary terms in the context of the Government's introduction of a badly implemented fee system in Employment Tribunals that had the effect of curtailing access to justice for many potential litigants. Another dimension of access to justice is the function of a justice system: although a lot of what lawyers and judges do is (one way or another) focussed on outcomes, the process itself can be just as important. Many of us have will have been involved in litigation where the most

[12] https://arstechnica.com/cars/2018/06/why-emergency-braking-systems-sometimes-hit-parked-cars-and-lane-dividers/

[13] *R (on the application of UNISON) v Lord Chancellor* [2017] UKSC 51

important thing the judge did was to listen. People often need to be listened to whatever the ultimate outcome of their dispute. The knowledge that they have been given a fair hearing is a key part of accepting an adverse decision. This is most obviously so in areas such as crime and family law, but it commercial actors too may benefit from participation in a non-automated system of legal dispute resolution.

CONCLUSION

The widespread uptake of predictive coding as an acceptable means of complying with the obligation to disclose documents in litigation shows how readily the litigation and arbitration processes can adapt to technological innovation. That, combined with the large amounts of money at stake in commercial litigation, provides fertile ground for the use of artificial intelligence both by both individual actors and the wider system. However, while the first examples of artificial intelligence in the law moving beyond academic research have already been observed, a mixture of technological and wider challenges exist.