FEATURE ARTICLE

The Genie and the Lamp: How Can Artificial Intelligence Help Us Find New Case Law?

Abstract: Based on a presentation given at the BIALL Annual Conference in July 2022, this article by Paul Magrath provides an overview of how technology including artificial intelligence (AI) is transforming legal practice and the conduct of litigation, followed by more detailed consideration by way of a case study of ICLR's development of its AI-driven search tool, Case Genie. The article examines the problems that it was designed to solve, particularly the legal researcher's anxiety over 'unknown unknowns', and the options for further development of the technology and its application in other areas.

Keywords: artificial intelligence; case law; law reporting; legal research

INTRODUCTION

There's been a lot of hype around artificial intelligence generally, and about AI in law in particular. Much of the mystique around it stems from simple ignorance. As the saying goes, 'It's only AI when you don't know how it works; once it works, it's just software.'

In a nutshell, what AI can do is replicate the work of human intelligence in humdrum, repetitive or time-consuming tasks, in support of human decision making rather than instead of it. Typically, this will be tasks such as searching through vast swathes of data, automating processes such as contract drafting, sifting and selecting documents for e-discovery, or modelling outcomes from different courses of action.

Less well explored is the potential of AI in developing tools for case law research. That is one of the reasons why, in 2019, the Incorporated Council of Law Reporting for England & Wales (ICLR) set up its own research and development lab, ICLR&D. Its purpose was to explore the possibilities of making better use of case law data, including ICLR's own vast hoard of law reports going back to 1865, and how it might be harnessed to serve the needs of ever more demanding users.

This article charts ICLR's subsequent journey in researching and then applying AI in the development of case law research tools, notably our Case Genie recommendation engine. All this is viewed in the context of other developments in the legal landscape in recent years, including the increasing technical demands of the legal professions and the modernisation of the administration of justice.

THE LEGAL LANDSCAPE IN 2022

The legal industry is not monolithic. Its needs for information are various and will never be satisfied by a 'one size fits all' solution. But what nearly all legal work shares is the need to use and analyse data, whether that be in the form of words and documents or figures and statistics. Data can be aggregated and analysed but doing this with human resources can be very time consuming and therefore expensive. What is often referred to as 'Lawtech' has for obvious reasons enjoyed a boom in recent years.

When thinking about case law research, we recognize that different sectors of the legal professions have different needs. For example, barristers will typically be researching cases for an appearance in court, or possibly for the drafting of an opinion. Judges may want to ensure that the decision they give, which relies heavily on materials provided by the advocates representing the parties, hasn't overlooked something such as to render their decision 'per incuriam'. Solicitors may require an up to date view of the law on a particular topic in order to advise a client as to a particular transaction or course of conduct. A librarian or personal support lawyer may need to prepare a briefing based on the latest available information. Academics will want to ensure that their teaching materials are up to date, and those engaged in research will want to ensure they, too, have access to the full range of information on the topic.

Although traditional methods of print-based research remain essential in some contexts (particularly in relation to older or more obscure sources), the majority of research is now done online. The last two decades have seen a rapid and accelerating development in both the volume and accessibility of legal materials online. But while the legal industry has adapted rapidly over the last three decades in its use of digital tools and materials, the same cannot be said for the courts – currently managed

by His Majesty's Courts and Tribunals Service (HMCTS) under the joint direction of the Ministry of Justice and the Judiciary. However, all that is now changing.

In 2014 the then Justice Secretary Chris Grayling announced a massive £1bn project to transform the outdated court system. That project, known as HMCTS Reform, was expected to take five years from 2015. It's been extended and overrun twice and is now expected to take more like eight years.

What does 'Reform' mean? I have written a fuller analysis of this project in a previous article[1] so I will merely summarise here. Reform means, first, the replacement of paper-based systems with digital systems. The days when one could wander down the corridors of court buildings and find boxes full of cardboard folders and lever arch files have now gone the way of vellum and quill pens. Court filings and bundles are now lodged and exchanged in digital form, online.

Secondly, court buildings are being rationalised. Smaller, uneconomic courts have been closed and merged into larger, often multi-jurisdictional court complexes, such as the massive new Civil Justice Centre in Manchester. Unfortunately, this has also meant a loss of local access to and scrutiny of justice, as many smaller local county and magistrates' courts have been closed, the buildings sold off, and cases transferred to the bigger metropolitan court centres.

Thirdly, one of the factors justifying the closure of local courts has been the increasing use of video links for witnesses giving evidence or defendants making bail applications from prison cells, reducing the need for travel to and physical attendance at court. With the onset of the Coronavirus pandemic, and emergency legislation[2] to permit even broader exceptions to the long-standing statutory prohibitions on filming and recording in court,[3] the occasional use of video evidence has morphed into a wholescale use of video conferencing technology (the so-called 'Zoom boom') to conduct entirely remote hearings.

The fourth and final aspect of Reform is the creation of an entirely online court. This is designed to deal with money claims and other relatively straightforward procedures. It is based on the successful model of British Columbia's Civil Resolution Tribunal. There is a three-stage process, the first stage of which is likely to be entirely automated; at the second stage, a case worker may intervene to suggest a mediated solution; only if that fails need the case progress to the third level, with a judge providing the determination.

When we talk about 'online courts' or 'online justice', it is important to maintain the distinction between a remote but live (synchronous) hearing conducted via an internet connection, such as a Zoom or Teams meeting, but still essentially conducted within the existing court structure; and the separate online court, with its own forms of application and rules of procedure, and with hearings conducted asynchronously, if at all.

AI AND LEGAL TECH: AN OVERVIEW

Let's begin with a couple of definitions.

Natural language processing (NLP) is the computerized analysis or 'parsing' of ordinary human speech or writing in order to identify and categorise the concepts, ideas and entities described in it, and to assign to them a value that the computer can then match up with or relate to values derived from other concepts, ideas and entities.

Machine learning (ML) is another. It is the process by which an AI model is 'trained' using existing data and thus 'learns' to recognize the different concepts, ideas and entities which it is being asked to detect during the parsing process of NLP. ML requires large volumes of existing data, which are fed through the NLP and then 'corrected' – in essence told whether it has guessed correctly or not – thus providing a sort of confirmation feedback loop.

Most lawtech AI products use NLP in some form or other in their operation, and will have benefited from ML in their development. A good example of this is a chat bot. Whether orally or in writing, the chat bot receives a query from a human in natural language which it must then parse to interpret what the human is asking and break it down into a form of data that the computer can then match up with data representing the relevant answer. At its most basic, the chat bot will simply recognize a word or phrase from a lexicon it has been given and match it up with an answer from a crib sheet. A more sophisticated chat bot might give the impression of actually understanding human speech, but it's an illusion.

An example of a legal chat bot that appeared to work well was developed by a British student who was fed up getting parking tickets in London. When he moved to Silicon Valley, Joshua Browder developed an app called DoNotPay which was trained to convert the lay person's description of the problem or complaint into legal language and respond with the correct language to use in order successfully to contest the ticket. It has since been expanded to include other services as well and is available on subscription in the UK and USA. It won an award from the American Bar Association for 'increasing legal access'.[4]

Such products are sometimes referred to as a 'robot lawyer' but there is no sense in which a chat bot could actually replace a fully trained human lawyer. However, they might replace some of the functions of a paralegal. The initial stage of the online court described above might use a similar process, by way of an online form or decision tree, to "triage" a lay litigant's claim and assess its suitability for online resolution. The challenge is to anticipate the tendency of humans to misdiagnose and therefore present their problems in a confusing or inconsistent way. That's why it's likely to work best by asking simple questions and using the answer to determine the next question.

But another use of NLP is in the analysis of different draft or legacy wordings of contracts in order to detect

errors, omissions, ambiguities or conflicts. The creation of new contracts can even be automated, using AI. (This feeds into or is perhaps an aspect of other developments such as 'smart contracts', in which aspects of the drafting, conditions or operation of a contract are to a greater or lesser extent automated or transactions recorded using blockchain or distributed ledger technology. Such contracts can be in code or natural language or a mixture of both.)

NLP is also used in sifting and selecting relevant materials for e-discovery. That is the process of identifying, preserving and, where required to, serving copies on the other parties to litigation of, documents and other materials stored electronically.

Compliance is another area in which AI is being used. For example, the UK Intellectual Property Office recently launched a new AI screening service for trade mark applications which checks them against the existing register and assesses them for compliance and coherence.

Some AI models purport to predict future behaviour based on data about previous behaviour. In the case of criminal suspects, this could be to assess whether and how likely they are to reoffend while on bail or present a flight risk, or to re-offend when sentenced. But such models have been shown to demonstrate bias by reason of bias in the underlying data. Typically, the data might be skewed because it relies on proxies, such as neighbourhood crime statistics, or racial profiling. Evidence of such bias has been found in a widely used commercial tool used in the United States to measure the risk of recidivism. Researchers found that black defendants assessed using Northpointe's tool COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk.[5]

In the case of judges, the prediction model is based on their earlier decisions. Some studies have purported to show unconscious bias, while others have shown a tendency for judges to act more harshly in cases when tired or hungry, for example towards the end of the session, and less so when recently refreshed. The revelation of bias or inconsistency in judicial behaviour can be uncomfortable. But it's also prejudicial against a judge to assume that they will always decide in a particular way. It's a denial of their free will. This issue has grown in importance in light of a recent debate in France where legislators actually decided to outlaw judicial analytics, as being likely to undermine the dignity and independence of the judiciary. However, the practice is common in the United States, where the volume of data available makes the analysis rather more likely to be reliable. It's important to note, however, that such analytics are based on outcomes and rely on structured data such as court filings or dockets – to match up the nature of the claim, the identity of the judge or court, and the outcome reached – rather than the loosely structured wording of court judgments (or opinions).

Would a cyber judge be better? The Washington Post recently reported that an AI model has been developed that replicates how the late Justice Ruth Bader Ginsberg would respond to a particular legal question.[6] It was trained on all her opinions and writings and various public statements she'd made over her life. If they can do that with a former US Supreme Court justice, you might think they could do the same with, say, Lord Sumption or Lady Hale. We might even solve the judicial recruitment crisis.

Real judges can breathe a sigh of relief, however. Paul Schiff Berman, a law professor who had once clerked for RBG, fed the bot a question he'd always meant to ask the legendary judge, and was not impressed with the answer. It didn't directly answer the question and its reply implied that Ginsburg did not believe in the judicial concept of deference, which was not true, Berman said. He also noted that the model did a poor job in replicating her unique speaking and writing style.

ICLR'S JOURNEY

As a traditional print publisher, ICLR was slow to adapt to the internet. When I took over my current role in 2008, we still thought we were pretty clever allowing people to read our cases on CD-ROMs. All that changed with the launch of ICLR online in 2011. As print subscriptions have fallen, and online subscriptions have risen, ICLR has continued to develop the ideal platform on which to find them and use our law reports and associated content, such as unreported judgments and legislation.

In 2019 we set up our own research and development lab, ICLR&D. We undertook a programme of research to see what else we could do, both in terms of content, and looking behind that, in terms of data. We looked at what other providers were doing, and at what some of the larger firms were working on in their own labs. We came up with a number of ideas, but the main one was to use natural language processing to create a way of analysing, classifying and then comparing cases at a fundamental level.

We began by developing a prototype, which we called the Blackstone Library. We used machine learning techniques, to train a piece of open source software on our existing massive corpus of case law. We taught it to recognise and pick out legal terms, entities and concepts. How did we do this?

1. Firstly, we assembled the training dataset. We took the text of all our cases and divided it into individual sentences, then divided the sentences into individual words.
2. Then, having 'tokenised' the words, we parsed them for parts of speech and dependency.
3. The next stage was to train the Named Entity Recogniser (NER) to pick out such entities as 'name', 'citation', 'statute', 'court', 'judge', etc.

4. We also trained the text categoriser to pick out such concepts as 'axiom', 'conclusion', 'issue', 'legal test', etc.

The Blackstone Library was the brainchild of our then Head of Research, Daniel Hoadley – who has since left ICLR to become Head of Litigation Data at Mishcon de Reya. He in turn had been inspired by seeing something similar being done by platforms such as CaseText in the United States, which had recently launched its 'brief analysis' tool, CARA AI. Daniel built Blackstone using spaCy and Python. He published his code on Github and described it all on the ICLR&D blog.[7]

Having built a proof of concept, the next stage was to find a use for it and develop it at scale.

LEGAL RESEARCH: 'UNKNOWN UNKNOWNS'

We knew from our research that a key problem for lawyers was 'unknown unknowns'. As that great exponent of military epistemology, Donald Rumsfeld, famously observed:

1. There are *known knowns* – which in case law research means cases that we already know about. We can use a name or citation to find them on the shelf or online.
2. There are *known unknowns* – which in case law research means cases that we know must exist. We may not know their names or citations. But we can find them. They will be referred to in other cases or we can look them up.
3. But what don't we know? There are cases we don't know about because they are new cases, or because they may come from an area of law we might not think of looking at. They might be relevant but we aren't even aware that we don't know about them. That's an *unknown unknown*.

Traditionally, the only way of knowing about unknown unknowns was by relying on someone drawing them to your attention, usually by reporting or commenting on them, or tweeting about them, or at least indexing them with some form of topic classification. Yet for a variety of reasons the thing you're looking for might not be the sort of thing that gets reported.

The problem – the need to discover unknown unknowns – usually arises in the context of some task. For a barrister, that might be preparing a case for trial. For a solicitor it might be checking what the latest law on a particular topic is. For a student it might writing an essay or preparing for a moot. In all these scenarios, there is often an *existing piece of text* in which the problem is described. A classic example would be a barrister's skeleton argument. But it could be the first draft of an essay or article, or an internal research briefing.

Using something like the Blackstone Library we could analyse it, draw out the concepts and ideas (and existing cases) found in the document, and then find similar or related cases from the vast hoard of ICLR published judgments. We took this idea to our developers, 67 Bricks in Oxford, who built a real live working version of what had till now only been an experiment. The result was Case Genie.

How does it work?

First, the user must upload a piece of text for analysis. It works best if the input text is focused on a single issue or legal topic. The more focussed the input text, the more focussed the results will be.

Text can be uploaded as an entire file, direct from the user's computer, in Word, PDF or text format, or by the user simply pasting in selected text into a box on the Case Genie page. The input text is immediately encrypted, because we know that lawyers are concerned about confidentiality.

The text is then passed through a pipeline of processes, during which it is tokenised – broken down into sentences and individual words, and parsed for parts of speech and dependency. Entities such as case names, citations, courts, and legal concepts and ideas are identified and isolated. We then use a classification library program called fastText to look up and assign to each word vectors from a model which we created using ICLR's own existing content, and a library called FAISS (Facebook's library for similarity search for very large datasets) to manage the similarity comparison process.

What do we mean by vectors? Think of a vector as a location plotted on a two-dimensional graph. Each vector consists of a value and a direction. Now expand the number of dimensions to 600. This is where it gets mysterious, at least for us muggles who don't normally think in terms of conceptual mathematics. But with those 600 dimensions, what it means is that each word is assigned an "embedding" comprising 600 vectors. Those vector readings are then used to compare the input text or document with ICLR's existing corpus, to find and suggest the most closely matching cases in terms of subject matter, in both fact and law.

Once the analysis is complete, the user is presented with a list of results of up to 50 suggested cases. These results can be filtered down or searched within, in the same way as in any other search results on the ICLR.4 platform. The difference is that the search has been "primed" by confining the results to those suggested by Case Genie.

At this point the user has a number of options. They might first wish to check the cases they have already included in the input text. There is an option to display these on a separate browser tab or page. These are your *known knowns*, if you like. Case Genie has looked them up in the ICLR citator index and checked their status. Have they been cited by a later case, perhaps even overruled? All the citations found are listed, along with any other information, such as parallel citations, and an indication of their status derived from the ICLR citator index.

Another option, going back to the main results, is to add in the linked or related cases – cases which have cited, or been cited by, the existing or suggested cases. (The *known unknowns*, if you like.) If you think of case

considerations as a vast network of connections between cases, you can see how extensive such a process of linking might be. How many degrees of separation do you want? How significant must the consideration by the other case be? Even limiting it to, say, two degrees of separation or a particular level of consideration can still produce a vast number. But by adding these into the results, the search can be primed to ensure not only that cases on similar subject matter are recommended, but other relevant cases by which the law may have been developed are not overlooked.

What the research with early users told us was that, while Case Genie often produced results whose relevance was not immediately apparent, it can be particularly useful as a way of unearthing (a) cases from a different area of practice (where you might not have thought of looking) or (b) unreported cases which happen to contain something similar or relevant.

Discussing the project outside the legal community (for example at a tech conference) has also been interesting. People have suggested that the same technology could with minor adaptations be applied to other subject areas and disciplines, such as medicine, science or history. It might also be used for literary criticism, or at least to identify similarities and connections between different literary texts.

But it has its limitations. It's not a calculator. It isn't a question of putting in two plus two and expecting to get four every time. Because of the way it works, as a closed or 'black box' system, it is not possible to explain *why* any particular case result has been suggested. Hence the hint of mystery in the name. It is not a robot or cyber-librarian. Despite being shortlisted for two awards – for Most Innovative Use of AI/ML in Computing magazine's AI and Machine Learning Awards 2022, and for innovation in publishing in the ALPSP (Association of Learned and Professional Society Publishers) awards 2022 – we are still very conscious that there is further development to be done. We welcome feedback and we want to make it work better and do other things.

Footnotes

[1] Paul Magrath, 'Transparency, Data Protection and the Law Courts of the Future' (2018) 18 *Legal Information Management* 70–75.

[2] The Coronavirus Act 2020, sections 53–57 provided for temporary amendments to other legislation, as set out in Schedules 23–27, including new the Courts Act 2003, into which was inserted new sections 85A 'Enabling the public to see and hear proceedings', 85B concerned with 'Offences of recording or transmission in relation to broadcasting' and 85C with 'Offences of recording or transmitting participation through live link'. Those amending provisions have now been re-enacted on a permanent basis in the Police, Crime, Sentencing and Courts Act 2022.

[3] There are prohibitions against filming under the Criminal Justice Act 1925, s 41 and against audio recording under the Contempt of Court Act 1981, s 9.

[4] Would You Let a Robot Lawyer Defend You? (BBC, 15 August 2021) https://www.bbc.co.uk/news/business-58158820

[5] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, 'Machine Bias' (*ProPublica*, 23 May 2016) https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[6] Pranshu Verma, 'This AI model tries to re-create the mind of Ruth Bader Ginsburg', (*Washington Post*, 14 June 2022) https://www.washingtonpost.com/technology/2022/06/14/ruth-bader-ginsburg-ai/

[7] 'Open Source Natural Language Processing for Legal Texts', ICLR&D/Blackstone https://research.iclr.co.uk/blackstone

Biography

Paul Magrath is Head of Product Development at ICLR and editor of their blog, where he writes the Weekly Notes roundup of legal news and commentary. He has been reporting and writing about case law in a variety of publications for most of his adult life. He is also a founding member and trustee of the Transparency Project, whose objects include the promotion of open justice and of public legal education, particularly in relation to family law. He is the co-author, with Julie Doughty and Lucy Reed, *Transparency in the Family Courts – Publicity and Privacy in Practice* (Bloomsbury Professional 2018).