AI Decision-Making and the Courts

A guide for Judges, Tribunal Members and Court Administrators

Felicity Bell / Lyria Bennett Moses / Michael Legg / Jake Silove / Monika Zalnieriute

Joint research project between the Australasian Institute of Judicial Administration (AIJA) and UNSW Faculty of Law and Justice, through the Allens Hub for Technology Law and Innovation and the Law Society of NSW Future of Law and Innovation in the Profession (FLIP) research stream

AIJA
The Australasian Institute of Judicial Administration Inc.

The Secretariat
The Australasian Institute of Judicial Administration Incorporated
Level 12, 170 Phillip Street
Sydney New South Wales 2000
Australia

Telephone: (61 2) 8099 2611
Website: www.aija.org.au

The Australasian Institute of Judicial Administration Incorporated ("AIJA") is an incorporated association affiliated with Monash University. Its main functions are the conduct of professional skills courses and seminars for judicial officers and others involved in the administration of the justice system, research into various aspects of judicial administration and the collection and dissemination of information on judicial administration. Its members include judges, magistrates, legal practitioners, court administrators, academic lawyers and other individuals and organisations interested in improving the operation of the justice system.

The AIJA Secretariat, which has been in operation since February 1987, is funded substantially on a composite government funding basis through the Council of Attorneys-General (CAG).

© The Australasian Institute of Judicial Administration Incorporated

Published June 2022

ISBN: 978-1-875527-60-1

Foreword

In this technological era, Artificial Intelligence (AI), although still in its infancy, is slowly being introduced across all jurisdictions. However, it is evident there is a lack of real understanding across the legal profession of how these systems operate to aid the judiciary and tribunals. As there are strong indications that such tools will be increasingly deployed, it is important for the judiciary, tribunal members, and court administrators to be made aware of the latest developments within this field.

This report seeks to be a guide for courts and tribunals presenting a brief overview of various AI and automated decision-making tools alongside the challenges and opportunities they present. Through survey analysis and virtual roundtable discussions with AIJA members, the research team was able to identify the primary interest areas of judges, tribunal members and court administrators raising important questions which should be asked when using AI. The report provides different areas where AI can be utilised as well as how such tools can impact core judicial values of open justice, accountability and equality before the law, procedural fairness, access to justice, and efficiency.

On behalf of the AIJA, I would like to thank the authors of the study, Professor Lyria Bennett Moses (Director of Allens Hub for Technology, Law and Innovation, Faculty of Law & Justice, UNSW), Dr Monika Zalnieriute (Senior Lecturer, Faculty of Law & Justice, UNSW), Professor Michael Legg (Director of the Law Society of NSW Future of Law and Innovation in the Profession (FLIP) research stream, Faculty of Law & Justice, UNSW), Dr Felicity Bell (Research Fellow, FLIP research stream, Faculty of Law & Justice, UNSW) and Jake Silove (Lawyer, Australian Government Solicitor).

The report provides useful points which should be considered by court administrators and managers when adopting certain AI tools. The authors have also helpfully identified the significant impact AI tools have played in the administration of justice and the judicial system. Such work is hoped to provide a fuller picture in utilising tools to promote efficiency and justice within the court system.

The AIJA is proud to continue supporting research into the administration of justice, and to continue our role in judicial education.

The Honourable Justice Jenny Blokland
Supreme Court of the Northern Territory,
President, Australasian Institute of Judicial Administration

Preface

The Australian Institute of Judicial Administration (AIJA) engaged researchers at UNSW Law and Justice to prepare a guide for judges, tribunal members and court administrators in the Asia-Pacific region on artificial intelligence (AI) in the courtroom. The guide addresses:

- Key challenges and opportunities that AI tools present for courts and decision-makers;
- Different techniques falling under the umbrella of AI, their affordances and limitations;
- Examples of different areas where these techniques have been used in courts, both regionally and globally, together with a discussion of important issues arising in those contexts; and
- Interaction between such uses and core judicial values.

To determine the scope of the guide, the UNSW research team conducted a survey of AIJA members to establish the areas of greatest interest. A copy of the survey instrument is in Appendix 1. Following preparation of the draft guide, a virtual round table meeting took place in November 2021 to seek feedback from AIJA members. This helped the UNSW research team to improve navigability so that judges, tribunal members and court and tribunal administrators can use the guide as a tool to ask important questions when considering the use of AI to perform particular tasks. While the document may also be useful for those appearing in courts and tribunals, litigant support groups, and policymakers, they are not the primary audience.

Table of Contents

1. Introduction

Artificial intelligence (AI) systems pervade modern life and are already being used in courts, both in their administration and to support decision-making, and by the legal profession. An understanding of AI is becoming increasingly important for judges, tribunal members and court administrators. It is also important in the context of statutory interpretation.[1]

This guide sets out the key challenges and opportunities that AI and automated decision-making presents for courts and tribunals. It draws on legislation, case law and rules in jurisdictions including the United States of America, England and Wales, and the European Union. The guide is not intended to provide an exhaustive analysis of emerging technologies, AI tools and the courtroom. Instead, it overviews some of the ways in which AI may be incorporated into domestic courtrooms and analyses some associated benefits and risks. Given that technology continues to evolve, the guide starts with the function and purpose of the technology and its impact on foundational values which underpin the judicial system.

The following section introduces common AI terms and techniques, ranging from older tools, such as expert systems, to more recent developments in machine learning. Section 3 then outlines common areas of AI use by the courts, or by parties, lawyers and legislators where that impacts the courts. Section 4 discusses how AI tools, when used in the courtroom, impact on the core judicial values of open justice, accountability, impartiality and equality before the law, procedural fairness, access to justice and efficiency. These values interact and often overlap with one another, including in the context of AI tools. Yet, they are useful guiding points for understanding how AI systems have the capacity to impact on the courts and judiciary.

1 For example, in the recent case of *Thaler v Commissioner of Patents* [2021] FCA 879. Beach J found that an 'artificial intelligence system or device' can be an inventor of a patent, opining that '[w]e are both created and create. Why cannot our own creations also create?' at [15]. This decision is under appeal.

2. Common AI Terms and Tools

2.1 Artificial Intelligence

AI is a broad umbrella term with no single meaning. Originating in the 1950s, it is used loosely to refer to many different areas of computer science, such as machine learning, computer vision, natural language processing, speech recognition, robotics, expert systems, and planning and optimisation.[2] The term 'AI' commonly features in social and cultural debates in relation to ethics, risks, regulations, human rights and the future of humanity. AI is often understood as machines displaying human-like intelligence,[3] yet that is not exactly accurate. Computers can perform various functions, but it does not mean they are 'intelligent' or self-aware about their operation. It has also been argued that AI is not 'artificial' because it is made from natural and human resources and depends on wider political and social structures.[4] The term 'complementary' rather than 'artificial' intelligence thus might be more suitable to describe the phenomenon if our goal is to create systems that solve problems that are difficult for humans rather than to duplicate human intelligence.[5]

The OECD defines an 'AI system' as 'a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments'. It notes that 'AI systems are designed to operate with varying levels of autonomy'.[6] Other bodies define AI differently. Due to the lack of a robust definition in Australia, the Asia-Pacific or internationally, the meaning of the term AI is contextual and may be defined differently in legal instruments,[7] policy settings, or in contracts as part of a description of goods or services. Thus, legal requirements, contractual promises and dialogue that refer to AI should be understood and interpreted with reference to how the term is used in the specific context.

2.2 Expert Systems and Traditional Programming

Expert system refers to a 'first-generation' AI system, where AI relies on knowledge provided by a human expert in a domain, such as law, to make predictions, recommendations or decisions based on that knowledge. A process in the expert system can be automated using a series of explicitly programmed steps and so-called 'if...then...' rules to create a 'rule-set', which a computer can implement. These rules can be expressed visually in the form of a decision tree, where the available choices are referred to as 'nodes'. Figure 1 is an example of a decision tree which determines whether a person can vote in an election in a country in which the only requirements are that the person is over the age of 18 and a citizen of that country.

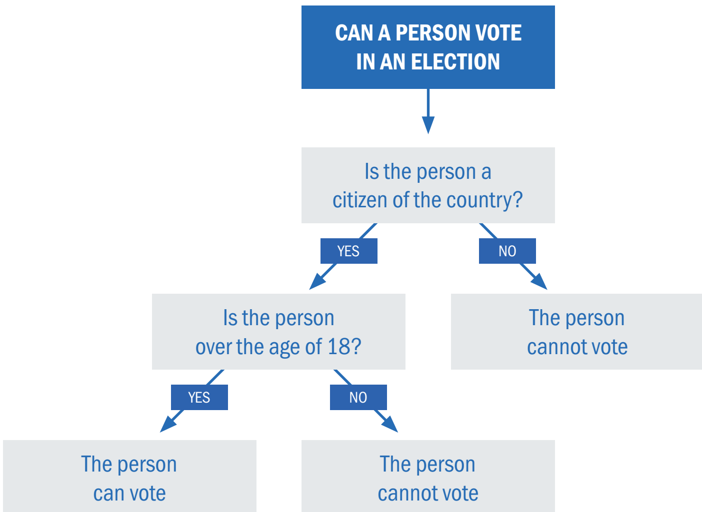


Figure 1: Example of decision tree

Figure 1 is an example of a 'binary' decision tree, as there are no more than two nodes stemming from each branch.

2 Julian Webb et al, *The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing* (ACOLA, July 2019) 14.
3 John McCarthy et al, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (31 August 1955) <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>.
4 Kate Crawford, *Atlas of AI* (Yale University Press, 2021) 7-9.
5 Roger Clarke, 'Why the World Wants Controls over Artificial Intelligence' (2019) 35(4) *Computer Law & Security Review* 423, 429–430.
6 OECD, 'Recommendation of the Council on Artificial Intelligence' (2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
7 For example, see the proposed Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts 2021/0106 (COD), 21 April 2021 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.

Legg and Bell explain that a 'rule-based or expert system decision tree has facts as its nodes, and rules for how the tree should be traversed. A user can be stepped through a series of questions to classify a problem and generate a pre-scripted solution'.[8]

Originally, writing rules in a language that a computer could implement required learning a programming language. The idea of an 'expert system' was that the rules could be crafted by a domain expert (for example, a lawyer) who did not themselves have programming skills. There are now a range of 'no-code' platforms that make it easy to 'program' a computer to follow a particular process or reach conclusions based on a series of rules. Examples of such platforms include Austlii's Datalex, Neota Logic, Realta Logic, Checkbox and Josef. Through these, and depending on the platform used, legal experts can use phrases, statements, arrows, drag-and-drop or drop-down menus or similar mechanisms to create a rule-set. Thus, a lawyer without programming skills can encode a decision tree, such as that shown above.

2.3 Automation

Automation refers to the degree that a system acts without human intervention or control in some domain. The concept is neutral as to the technical means through which automation is achieved. Automation and AI are hence overlapping, but distinct, concepts. The operation of the term 'automation' in practice can be illustrated with reference to the levels proposed by the Society of Automotive Engineers for automated vehicles (see Figure 2). The scale begins at zero (no automation, where the driver performs all driving tasks) through level 3 (conditional automation, where the driver is ready to take control when notified by the system) to level 5 (full automation under all conditions).[9]

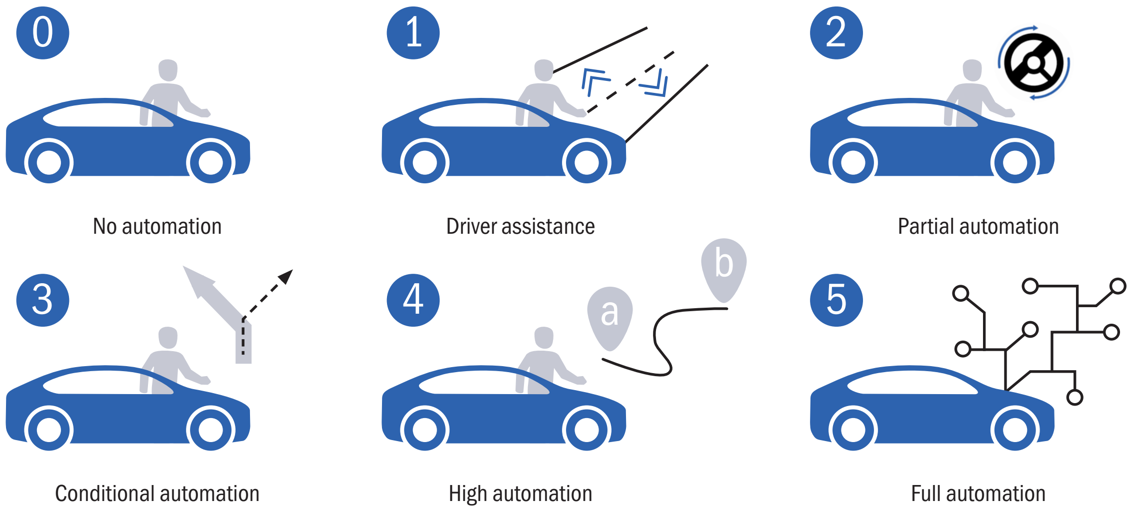


Figure 2: Levels of automation

An analogy can be drawn, albeit imperfectly, with automation in a courtroom. At the lower end of the scale is a courtroom in which all steps are considered and completed by trained individuals. This represents courtrooms prior to the advent and implementation of AI tools. Towards the middle of the scale is a courtroom reliant on some automated steps, such as automated e-filing, but which allows those automated steps to be amended or overridden by a human decision-maker. This represents many courtrooms in Australia and other jurisdictions. At the higher end of the scale is an entirely automated courtroom which generally, other than in exceptional circumstances, operates without any human decision-maker. As described in Chapter 3, such courts are being conceptualised and implemented in other jurisdictions.

8 Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart Publishing, 2020) 28.
9 'SAE International Releases Updated Visual Chart for Its "Levels of Driving Automation" Standard for Self-Driving Vehicles' (2018) <https://www.sae.org/site/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%9Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles>.

Automation can also describe the extent to which humans are involved in the system, using 'loop' metaphors:

- Human-in-the-loop: A human can change each output of a system; for example, a human must confirm a target before an automated weapons system fires. Confusingly, the same term is sometimes used to describe supervised machine learning (see below at Section 2.6) where data is labelled by a human.

- Human-on-the-loop: A human has oversight of a system but does not need to confirm an action; for example, a human can stop an automated weapons system from firing, but the system will otherwise automatically fire.

- Machine/AI/technology-in-the-loop: This language is used by some who argue that the human should be at the centre of a process, with technologies serving them.[10]

Other terminology that describes the relationship between humans and a system, particularly in the context of decision-making, is between a system that *makes* a decision and a system that *supports* a human decision-maker. For example, the output of a system might be framed as a *decision* that is implemented (by humans or by the system itself) or as a *recommendation* or *input* to a human-decision-maker, who may take other factors into account in making a decision. Sourdin uses the terminology of 'Judge AI' and 'supportive Judge AI' to articulate a similar distinction between AI that replaces a judge and AI that plays a role in decision-making processes.[11]

Where there is *no* external intervention, control or oversight of a system (by a human or by another system) once it is put into operation, the system can be described as autonomous. However, this does not imply that no person has *legal* responsibility for harm caused by such a system. Even an autonomous system has human designers, promoters, sellers, owners and users who might (depending on the circumstances) be legally accountable for its actions.

2.3 Bot

The term bot refers to an 'agent' that acts autonomously. Such an agent can be some lines of computer code, such as the automatic email replies that are sent out on behalf of employees on annual leave. Bots are used on social media platforms to generate social media content by automatically re-sharing content from other social media accounts. Some bots can be useful by automatically sharing certain information, such as statistics or scores from sporting matches. However, bots can also be used to spread disinformation, deceive or impersonate humans. In some jurisdictions, there are laws regulating social media bots – see, for example, the *Bolstering Online Transparency (BOT) Act* SB-1001 in California, USA.[12]

2.4 Rules as Code (RaC)

Rules as Code (RaC) is a public sector innovation, which involves a preparation of a machine-consumable version of some legislation. The term 'machine-consumable' implies that the rules are written in a way that they can be processed directly *as rules* by a computer. This can be done using a computer coding language or by using one of the platforms specifically built for this purpose. For example, Austlii's expert system platform Datalex allows legislation to be re-written in a machine-consumable format so that it can be queried through a chatbot. RaC is not appropriate for all legislation and is most useful for rules that involve a calculation, prescribe certain kinds of processes (such as a compliance process) or involve simple 'if-then' rules to determine matters such as eligibility for a benefit.[13]

As with other expert system techniques, machine-consumable rules can be written by lawyers or others without previous experience in computer coding. While RaC projects are conducted by the public service and do not directly involve courts, there may in future be implications for statutory interpretation and administrative decision-making. We therefore discuss RaC in section 3.10.

2.5 Algorithm

The concept of an 'algorithm' pre-dates the first computer and was created by a ninth century mathematician, Muhammad ibn Mūsā al-Khwārizmī. The term refers to set of non-ambiguous steps used to solve a class of problems or perform a class of computations, turning inputs into outputs. Thus, while computer programs are examples of algorithms, a primary school child doing long division is also using an algorithm. Despite its broad meaning, the term in popular discourse has recently come to be identified almost exclusively with machine learning algorithms.

10 Mireille Hildebrandt, 'Smart Technologies' (2020) 9(4) *Internet Policy Review* <https://policyreview.info/concepts/smart-technologies>.

11 Tania Sourdin, *Judges, Technology and Artificial Intelligence* (Elgar, 2021) 16.

12 *Bolstering Online Transparenc y (BOT) Act*, 7.3.6 Cal BPC §§ 17940-3. In that legislation, 'bot' is defined as 'an automated online account where all or substantially all of the actions or posts of that account are not the result of a person'.

13 New Zealand Government, *Better Rules for Government Discovery Report* (May 2018) <https://www.digital.govt.nz/>.

2.6 Machine Learning

Machine learning is 'second-generation' AI and the most well-known sub-field of AI research. Machine learning involves a model whose parameters are set through an algorithmic process to reflect data or specific experience. Machine learning has been incorporated in systems and software to solve a range of problems too complex for 'first-generation' AI systems or human decision-makers. The system is said to 'learn' because its performance improves as it processes data or experience. Yet, machine learning is not the same as human learning. A child only needs to be shown a few pictures of a cat to understand what a 'cat' is and identify other images that are cats. Computers can be trained to do the same classification exercise but will need a far larger training set. If the training set is too small and the number of features too large, then a model generated by a computer as to what a 'cat' is may 'overfit' the training data, rendering it too specific and therefore useless in classifying new data. The things that human learners and computer 'learners' are good at may thus be different.

The complexity of machine learning is best illustrated through the example of discovery of legal documents. Suppose we had a set of electronic documents and wished to work out which were discoverable in the context of particular litigation. This can be done in several ways (a more comprehensive description of actual practices can be found in section 3.1 Technology Assisted Review and Discovery; this example is intended to be illustrative only):

1. Mode 1, no automation: A human, usually a paralegal or junior lawyer, reads through the files and decide which documents are discoverable given a known set of parameters.

2. Mode 2, automation without machine learning: A set of fixed criteria (eg date range, list of words/phrases, file location, etc) is used to decide which documents are discoverable by having a computer system automatically search through the files for documents which contain the desired traits.

3. Mode 3, machine learning: A human decides (or 'labels') which of a sample ('training data') of the documents are discoverable. Criteria, such as, for example, date range, list of words/phrases and file location for determining discoverability can then be decided. Rather than specifying which criteria are necessary for discoverability, however, a machine learning system can be used to deduce these based on patterns among these elements in the human-labelled training data. The process may be able to identify patterns beyond those that might have been chosen using Mode 2. The trained model will use these patterns to categorise the remaining documents into those that are and are not likely to be discoverable.

The process in Mode 3 is called *supervised* machine learning because the system relies on training data that has been labelled (in this case, as discoverable or not discoverable). In *unsupervised* learning, patterns can be found in unlabelled data. For example, clusters of emails that use similar words and phrases could be identified. Such a system might identify that there are (say) three clusters of emails that tend to have similar language, length and format. The output itself will merely show that there are three clusters because the training data was not labelled; the system will not be able to ascribe any meaning to the distinction between the clusters. A person may look at the clusters later and conclude that there is a group of emails about organising meetings, a group of emails about sales figures and a group of emails about sales strategies. Such techniques may be used in an exploratory way when seeking to identify documents relevant to litigation.

In *semi-supervised* machine learning, only some of the training data is labelled. These methods are often used where labelling data is expensive and time-consuming but unlabelled data is easy to obtain. Varying the above email clustering example, one might label a small number of emails in the training set and use these to assign labels to emails that are, through analysing the labelled and unlabelled data together, in the same cluster.

In *reinforcement* learning, the learning occurs through a reward function that provides feedback while a system interacts with its environment so that the system can improve its strategy over time. For example, a system may learn to prefer moves in a game of chess that have, in the past, ultimately led to a victory. Reinforcement learning is often used where success depends on a series of steps (as in the chess example) rather than on making a series of discrete recommendations.

There are other contexts in which different machine learning approaches are important. One such approach is *continuous* learning, also known as *lifelong* or *continual* learning. Continuous learning occurs where the system continues to be trained – and thus to adapt and refine its performance – after it is already deployed in an operational setting. In continuous learning, the training and operational phases are thus not distinct. In Figure 3 Continuous learning, a machine learning model is initially trained using training data, perhaps from historic cases with a known outcome. After it is deployed in a real-world setting, the system is used on input data, yielding output data that has real world consequences, for example, making decisions that affect individuals. Data continues to be collected on what happens in those real cases and this information is used to further refine the machine learning model. In that way, the system will continue to learn while it is being used.

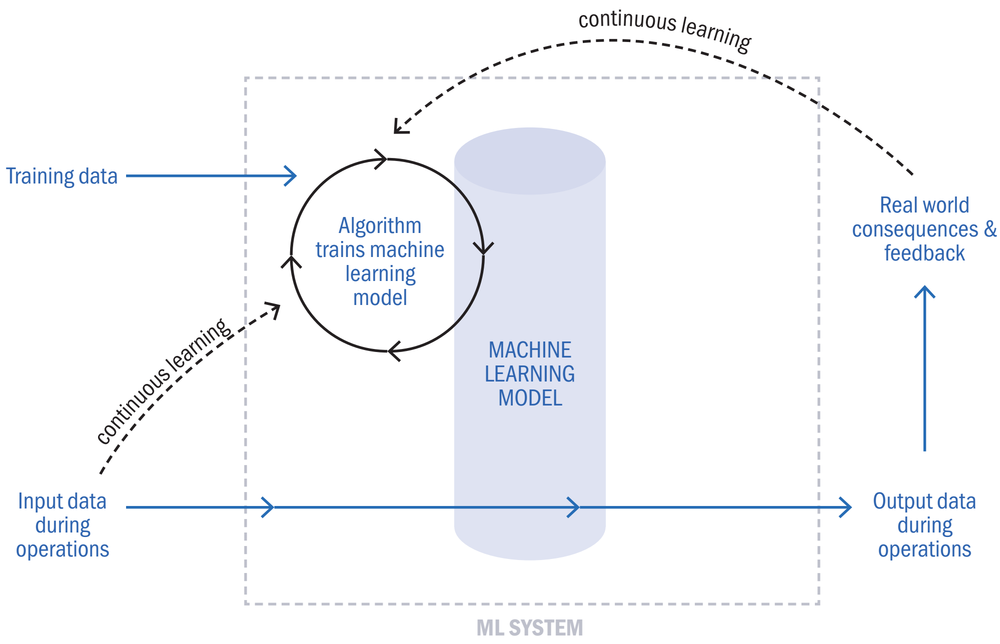


Figure 3: Continuous learning

A machine learning model is a mathematical construct. For example, a *linear* model assumes a simple relationship between two variables (say x and y) where we assume that $y = mx + b$ (where m is the gradient of the line and b is the point of intercept with the y axis). In machine learning, an algorithm is used to train the model. In the simple linear example, the system deduces the values of m and b that best fit the training data. Of course, the models used in machine learning are diverse and usually far more complex than a linear model. This section describes two examples of machine learning models. While *decision trees* (see section 2.2 Expert Systems and Traditional Programming) can be programmed into a computer, they can also be a very simple machine learning model. In such cases, the machine 'learns' the labelling and/or outputs associated with the tree's branches. As in all machine learning models, the output as only as good as the input data (see section 2.7 Garbage in – Garbage out). So, for example, while one could ask an expert to write a decision tree to identify those eligible to serve as President of the United States (natural born US citizen, at least 35 years old, resident in the US for at least 14 years), an attempt by an AI system to learn this from historic data could suggest alternative requirements such as being male, over 40, and not being from Alaska.

However, there are circumstances in which machine learning can be used constructively to build a decision tree. For example, Ruger et al used a decision tree machine learning model to predict the outcome of US Supreme Court decisions, achieving greater accuracy than human experts.[14]

Neural networks are another, much-discussed, example of a machine learning model. The model is inspired by the operation of the human brain (comprising neurons connected by synapses) but the analogy is imperfect and modern neural network techniques operate quite differently to a human brain. As an example, neural networks can be used to translate handwriting into a text document by recognising each letter or number (see Figure 4 Using a neural network to identify handwriting). The neural network will, using training data, make and weigh connections from the handwriting (the input layer) in the hidden, intermediate levels that represent components of letters and numbers which, when taken together, represent a particular letter or number (the output layer).

14 TW Ruger et al, 'The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking' (2004) 104(4) *Columbia Law Review* 1150.

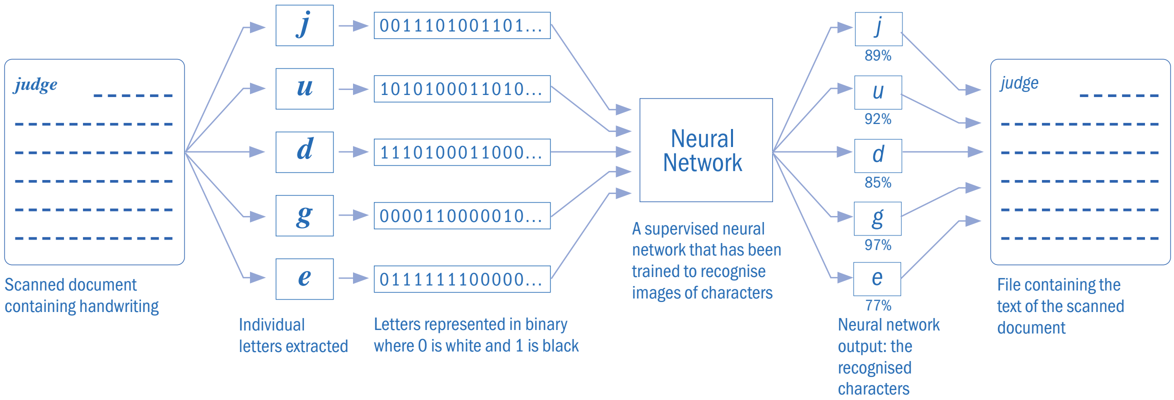


Figure 4 Using a neural network to identify handwriting

The basic unit in a neural network, called a 'neuron' or 'perceptron', has inputs and outputs. In the simplest scenario, the inputs of a 'layer' are the weighted outputs of perceptrons in a previous 'layer' and its outputs become weighted inputs for perceptrons in the next 'layer'. There are a range of different types of neural network, such as a feed forward neural network, a convolutional neural network and a recurrent neural network. While these differ significantly, there are common features, including the fact that the model itself is typically difficult to explain (in the sense of giving reasons comprehensible to a human as to why a particular output was generated). Where there are multiple hidden layers of perceptrons, the term 'deep learning' is often used (see Figure 5 A simple neural network).

A SIMPLE NEURAL NETWORK

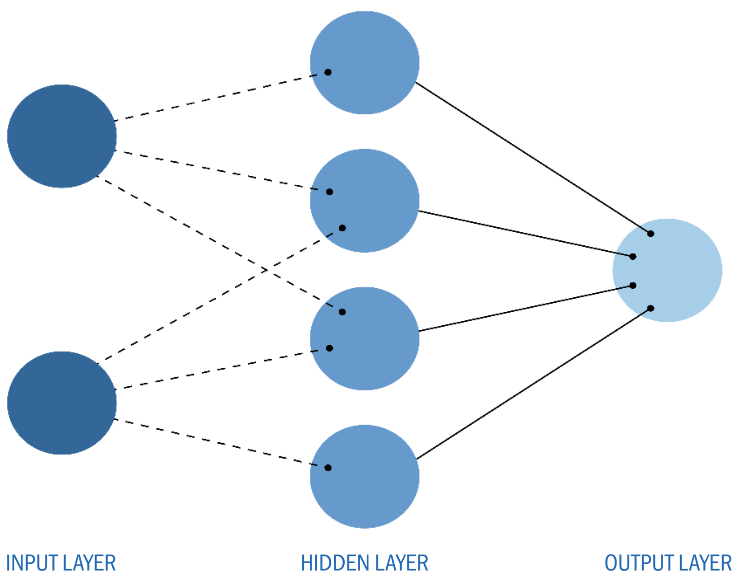


Figure 5: A simple neural network

Deep learning is used in tasks such as facial recognition. These methods rely on large datasets of faces to learn how to detect a face in an image, normalise the face (so that it can be compared to a face facing forwards), extract features that can be used to distinguish between faces, and then match the face to another (to confirm identity) or to a database of faces (to determine identity).

2.7 Garbage in – Garbage out

The reliance on data in machine learning means that the accuracy and reliability of the outputs generated will depend on the integrity and appropriateness of the training data that is used. If, for example, data collection was patchy so that it was systemically skewed, then the system will learn the same skew. When Amazon built a recruitment machine learning system that was trained based on data about its existing, largely male, workforce, the system 'learnt' to reject applications from women.[15] This problem is often neatly summarised as 'garbage in – garbage out'.

2.8 Bias

The term 'bias' is used in different ways within different disciplines. We look first at legal ideas about bias, then at technical bias.

Lawyers' concerns about bias do not relate to purely technical concepts, but rather to an unfair treatment of the kind that discrimination laws have traditionally dealt with. Such an unfair treatment can arise through the application of machine learning, either due to the model itself or through the data that the model is trained on. Eckhouse et al produced a useful framework through which to understand the ways in which bias may infiltrate an automated process (Figure 6):

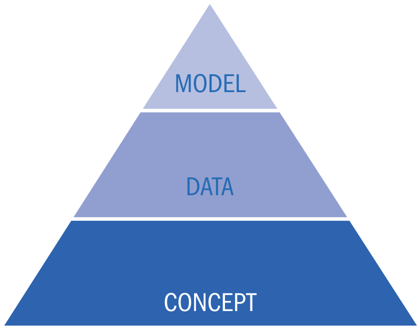


Figure 6: Layers of bias[16]

Eckhouse et al suggest that the fairness of each level is dependent on the ones beneath it. The top level in Figure 6, 'Model', involves the AI system itself and whether it contains any inherent discriminating functionality. The middle level, 'Data', incorporates the bias which can arise when the data used to train the automated system is itself infused with human bias (see section 3.4 Criminal Sentencing and Risk Assessment Tools). The lower level, 'Concept', relates to the underlying conceptual issues with the use of automated systems generally when determining the rights and interests of persons or parties to litigation. This foundational layer includes questions around the proper or fair way to make decisions about an individual based on aggregate or group data.

Bias can also have a technical meaning that is conceptually distinct from ideas of fairness or discrimination. In machine learning, 'inductive' bias (bias that arises through generalising from a sub-set of data) is inevitable. If we consider a set of data, there will be more than one rule that can explain that data. For example, the pattern 1, 2, 3 could be explained by counting, but it could also be explained by the rule 'if the number is less than 10, add 1; otherwise add 2'. One needs to make assumptions (for example, that the rule should be simple or that a relationship is linear), constituting inductive bias, to choose among the different things that might be learnt. When choosing machine learning models and algorithms, one is also choosing the nature of inductive bias and thus what kinds of errors are preferred. This is often done quite deliberately – for example, a government developing a machine learning system to classify threats to critical infrastructure may be more concerned about false negatives (threats classified as low when they are actually high) than false positives (threats classified as high when they could have been ignored).

When people with different disciplinary backgrounds discuss the term bias; for example, in the context of expert witness testimony, it is important to be clear on the sense in which the term is used. A data scientist may be talking about inductive bias whereas a lawyer may be concerned about fairness. The two concepts do intersect – for example, inductive bias that ignores 'outliers' may have negative impacts on minority groups. However, bias is not a purely algorithmic phenomenon, and a machine learning system may be unfair not because of bias introduced through a choice of model but rather through bias in the training data. AI systems may also be used to expose human biases which might otherwise be undetectable or unprovable.[17]

15 Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women', *Reuters* (10 October 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>.

16 Legg and Bell (n 8) 245.

17 Sharad Goel, Justin M Rao and Ravi Shroff, 'Personalized Risk Assessments in the Criminal Justice System' (2016) 106(5) *American Economic Review* 119.

2.9 Technological 'black box'

A technological 'black box' refers to human inability to grasp the inner workings of some technological systems. Even if humans can sometimes understand the inputs and outputs of a technological system, were they to view the inner workings of that system, they might find it incomprehensible. Accordingly, the person is unable to verify the integrity of the *process* used by the AI system to arrive at the output from the input. An explanation of connections in an artificial neural network is as unhelpful in understanding the system as is a neuron-by-neuron description of a human brain in understanding the reasons for a complex decision made by a human. This has led to interest in 'explainable' AI.

2.10 Explainable Artificial Intelligence

Explainable artificial intelligence (XAI) is a sub-discipline within AI which seeks to 'explain' AI and overcome the black box problem. Researchers in XAI focus on developing AI models that can be understood and interpreted by humans and on generating useable explanations of machine learning outputs. An example of an interpretable model is a decision tree – it is easy to understand how a decision tree operates to make decisions. 'Explanation' refers to numerous ways of exchanging information about a phenomenon, in this case the functionality of a model or the rationale and criteria for a decision, to different stakeholders.[18] An expert system can also generate explanations; it is possible to observe this in action by playing with some of the application examples on AustLII's Datalex system.[19]

The kind of explanation ought to vary depending on the context of use as well as the purpose of the explanation. For example, a consumer buying an automated vehicle will want to know about road testing and how different features work; they are unlikely to be interested in a live explanation of why the car adjusted slightly to the left on the highway. On the other hand, a system used in administrative decision-making should meet similar reason-giving requirements to a human making an administrative decision.[20] Similarly, a detailed explanation, constituting verification, will be required for a system determining the results of an election.[21] Explanations may also be useful when humans are working alongside machines so that they can better predict the behaviour of those machines. In some circumstances, explanations are required by law, as in the case of the EU *General Data Protection Regulation*.[22] The audiences of each of these explanations will also be different – with some having more technical understanding than others.

THINGS TO CONSIDER– Is an explanation sufficient to explain the operation of an AI system?

When considering an explanation offered for the outputs of an AI system, it may be helpful to ask the following questions:

- What criteria is the explanation required to meet? For example, is there a legal or contractual requirement to provide a particular kind of explanation?
- Does the explanation meet those criteria? Is it possible for *any* automated system to meet those criteria?
- Does the explanation concern the operation of the system as a whole or the rationale behind a particular output? Which is required or more appropriate in the relevant context?
- Is the explanation reliable? Is it possible that a system can generate an explanation that does not correspond to the internal logic of that system?
- Is the explanation comprehensible to the intended audience?
- Does the explanation address the things that the audience has a right to know, or might reasonably want to know, about the process?

18 Brent Mittelstadt, Chris Russell and Sandra Wachter, 'Explaining Explanations in AI' [2018] *arXiv* 1811.01439.
19 'DataLex', *AustLII Communities* <http://austlii.community/foswiki/DataLex/WebHome>.
20 Lyria Bennett Moses and Edward Santow, 'Accountability in the Age of Artificial Intelligence: A Right to Reasons' (2020) 94(11) *Australian Law Journal* 829.
21 Lyria Bennett Moses et al, 'No More Excuses: Automated Synthesis of Practical and Verifiable Vote-Counting Programs for Complex Voting Schemes' in *Electronic Voting: Second International Joint Conference* (2017) 66 <http://dx.doi.org/10.1007/978-3-319-68687-5_5>.
22 *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1, art 13(2)(f).

3. Areas of AI Use in Courts

AI systems are increasingly used in litigation and the courtroom in jurisdictions around the world, ranging from Australia, China, the United States and the United Kingdom to Estonia, Mexico and Brazil. Various AI systems are being built, tested and deployed in courts and tribunals globally, with new methods continually being developed. This section discusses examples to outline the main areas of implementation.

3.1 Technology Assisted Review and Discovery

Technology Assisted Review (TAR) is '[a] process for Prioritizing or Coding a Collection of Documents using a computerized system that harnesses human judgments of one or more Subject Matter Expert(s) on a smaller set of Documents and then extrapolates those judgments to the remaining Document Collection'.[23] A document is a discrete item of Electronically Stored Information (ESI) and a collection of documents is created by searching for or gathering documents that may be relevant to the issues in a dispute. The searching or gathering of documents will frequently utilise computers, but as explained briefly in the section above (see section 2.6 Machine Learning), it did not traditionally involve AI. TAR becomes useful when the volume of ESI is very large, such as discovery involving thousands or even millions of documents. As the volume and size of litigation continues to increase, the use of TAR in the discovery process is likely to expand.

TAR uses machine learning's capacity to identify patterns in textual data. Different forms of TAR exist: simple passive learning, simple active learning, continuous active learning and other systems.[24] Each of these are examples of supervised machine learning, as humans – preferably a lawyer familiar with the case – code documents and review (correcting where necessary) the AI system's categorisations. Human review is needed to 'teach' the software whether it has classified different documents correctly, and the method for teaching the software about which documents are relevant is referred to as a TAR 'protocol'.[25]

Starting with *simple passive learning*, the program is provided with a set of documents referred to as a training set. A lawyer reviewer codes the documents in the training set, labelling them (for example) as responsive or non-responsive. Using this information, the program applies this to other documents. Using the training set the software creates a model or classifier which 'can then predict the classifications of other documents'.[26]

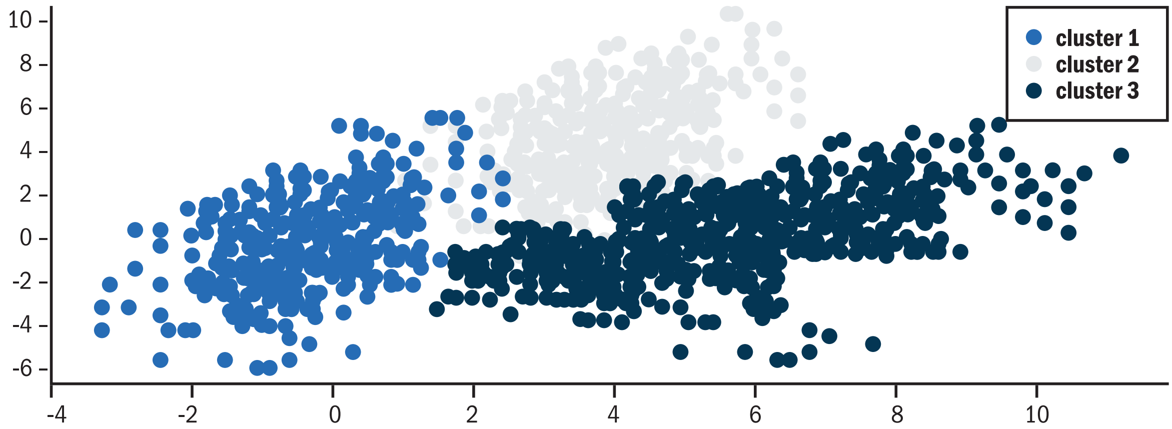
Simple active learning is where the software chooses some or most of the documents for training. The lawyer still needs to code the documents, but the software can identify the documents that will be most useful to it in developing its model or classifier. The software identifies documents for coding based on uncertainty sampling, i.e. the documents it is most uncertain about in relation to relevance.[27]

The TAR process described above, where there is a training set followed by several rounds of sampling and corrections, may be contrasted with an alternative approach called *continuous active learning*, or what has been called TAR 2.0. Here, the human review and the machine learning training process are combined; review and training occur simultaneously. Due to greater computing power, the system continuously analyses the entire document collection and ranks the population based on relevance. Human coding decisions are submitted to the system; the system re-ranks the documents, and then presents back to the human additional documents for review that it predicts as most likely relevant.[28]

Supervised machine learning:	simple passive learning simple active learning continuous active learning
Unsupervised machine learning:	clustering

23 Maura R Grossman and Gordon V Cormack, 'The Grossman-Cormack Glossary of Technology-Assisted Review' (2013) 7(1) *Federal Courts Law Review* 31.
24 See, Supreme Court of Victoria, *Practice Note SC GEN 5: Technology in Civil Litigation* (First Revision, 29 June 2018) 8.9.
25 Legg and Bell (n 8) 112.
26 Shannon Brown, 'Peeking Inside the Black Box: A Preliminary Survey of Technology Assisted Review (TAR) and Predictive Coding Algorithms for Ediscovery' (2016) 21 *Suffolk Journal of Trial & Appellate Advocacy* 237.
27 Grossman and Cormack (n 23); Jason Baron, Michael Berman and Ralph Losey, *Perspectives on Predictive Coding and Other Advanced Search Methods for the Legal Practitioner* (American Bar Association, 2016).
28 *Technology Assisted Review Guidelines* (Bolch Judicial Institute, Duke Law, 2019) 5; *ViiV Healthcare Company v Gilead Sciences Pty Limited (No 2)* [2020] FCA 1455, [135]-[138].

Another form of TAR is *clustering*, in which documents are segregated into categories or groups so that the documents in any group are more similar to one another than to those in other groups. Clustering methods measure the similarity of documents by using a geometric distance calculation and then cluster documents that are of geometrically similar distance. The system selects representative documents as the anchors for each cluster and then measures the distance of all other documents to the representative documents to group documents with a similar distance measure in similar clusters. Clustering involves no human intervention and is a form of unsupervised machine learning.[29]



The above descriptions are of generic approaches to the operation of TAR. However, TAR is offered in a competitive marketplace where TAR providers compete based on functionality.[30] As a result, the underlying methods employed, and the operation of, the program will vary. A TAR provider should have some form of written explanation as to how their product functions which may be more or less confidential.[31]

Courts in the United States, Ireland, England and Wales, and Australia have approved the use of TAR in the litigation process.[32] In *McConnell Dowell Constructors v Santam* (2016) 51 VR 421; [2016] VSC 734, Vickery J explained a simple passive learning form of TAR by reference to *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch).

In *Da Silva Moore v Publicis Groupe* 287 FRD 182 (SDNY, 2012), United States Magistrate Judge Peck observed that he was 'less interested in the science behind the "black box" of the vendor's software than in whether it produced responsive documents with reasonably high recall and high precision'.[33] His Honour was alluding to the fact that it was not necessary completely to understand how the AI system used in TAR functioned because its effectiveness could be assessed by reference to measures from the field of information retrieval: recall and precision. Precision is how useful the search results are, and recall is how complete the results are.

$$\textbf{Precision} = \frac{\text{Total number of documents retrieved that are relevant}}{\text{Total number of documents that are retrieved}}$$

$$\textbf{Recall} = \frac{\text{Total number of documents retrieved that are relevant}}{\text{Total number of relevant documents in the database}}$$

To provide an example, suppose TAR software retrieves 300 pages from a collection of documents and that only 200 of those pages were relevant while failing to return 400 additional relevant pages, its precision is 200/300 = 2/3 while its recall is 200/600 = 1/3. Precision is important because it means that only relevant documents are subject to manual review at the end of the TAR process and therefore costs are minimised. Recall is also important

29 Legg and Bell (n 8) 113-114.
30 Each year Andrew Haslam publishes a guide which sets out a list of suppliers and software in the UK market, including by way of example, NUIX, Relativity and Ringtail. See Andrew Haslam, 'The EDisclosure Systems Buyers Guide – 2021 Edition', *ComplexDiscovery* (4 June 2021) <https://complexdiscovery.com/the-edisclosure-systems-buyers-guide-2021-edition-andrew-haslam/>.
31 See eg Josh Borders, Brandon Gauthier and Elise Tropiano, 'White Paper: Active Learning in Relativity Assisted Review' [2018] *Relativity* <https://resources.relativity.com/white-paper-active-learning-in-technology-assisted-review.html>.
32 *Da Silva Moore v Publicis Groupe* 287 FRD 182 (SDNY 2012); *Irish Bank Resolution Corporation Limited v Quinn* [2015] IEHC 175; *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch); *McConnell Dowell Constructors v Santam* [2016] VSC 734; *Parbery v QNI Metals Pty Ltd* [2018] QSC 276.
33 *Da Silva Moore v Publicis Groupe* 287 FRD 182, 184 (SDNY 2012).

because it demonstrates compliance with the orders for discovery to find and produce the relevant documents.[34] As the total number of relevant documents in the database is unlikely to be known because of the high volume of documents, the recall of TAR can be compared with the recall achieved by a human reviewer coding a random sample of documents from the database.

While discovery is primarily the responsibility of the parties and their lawyers, the judiciary needs a familiarity with TAR where disputes arise, such as whether to use keywords to identify relevant documents or TAR,[35] in choosing between types of TAR and in addressing disagreements as to statistical parameters.[36]

Summary

TAR uses machine learning to review and classify high volumes of electronic documents. Its main use is in litigation to undertake large scale discovery. Parties will need to agree on key aspects of TAR such as the degree of accuracy that will be required.

3.2 Automated Online Dispute Resolution

Online dispute resolution (ODR) consists of online alternative dispute resolution (OADR) and online courts. OADR is dispute resolution outside the courts, which originally emerged in the mid-1990s as an adjunct to various forms of alternative dispute resolution (ADR) and as a response to disputes arising from the expansion of ecommerce.[37] As a result, it focussed on using technology to resolve customer complaints and sought to support negotiation, mediation and arbitration. Today it may go further and give rise to innovative ways to resolve disputes beyond the traditional categories of ADR. OADR may be privately run or state-sponsored, such as when it forms part of a consumer redress scheme. It may be synchronous (the participants are all present at the same time) or asynchronous (the participants engage with the process at different times), or a combination at different steps in the process. In contrast, online courts form part of the justice system and are therefore subject to institutional norms and legal requirements derived from the nature of the judicial function.[38]

Alternative Dispute Resolution – ADR

Online Dispute Resolution – ODR – may be:

- External to courts (Online Alternative Dispute Resolution – OADR)
- Within the courts

ODR may use a range of technologies, such as internet portals, email and audio-video conferencing facilities but, in relation to artificial intelligence, ODR typically employs the expert system, or decision tree analysis (explained in 2.2 Expert Systems and Traditional Programming). In the context of dispute resolution, human experts determine the questions that a citizen/client needs be asked to generate the information to determine how to proceed when faced with a particular legal problem. The aim is to structure the questions in a logical and user-friendly manner in order to identify the problem and then posit next steps. A successful expert system does not just accurately set out the necessary questions and information to provide, but also expresses it in an understandable manner for the non-lawyer user and provides a user-friendly interface.

An ODR process typically has three steps:[39]

1. Problem identification and provision of information.
2. Facilitation of voluntary forms of dispute resolution such as negotiation between the parties and third-party facilitated ADR such as mediation.
3. If step 2 is unsuccessful, preparation for initiating the steps to commence court proceedings.

34 Legg and Bell (n 8) 114.
35 See eg In re Mercedes-Benz Emissions Litigation (Case No.: 2:16-cv-881 (KM) (ESK) [D.N.J. Jan. 8, 2020]).
36 *McConnell Dowell Constructors (Aust) Pty Ltd v Santam Ltd (No 2)* [2017] VSC 640.
37 For example, eBay, PayPal and Alibaba.
38 Michael Legg, 'The Future of Dispute Resolution: Online ADR and Online Courts' (2016) 27 *Australasian Dispute Resolution Journal* 277; Richard Susskind, *Online Courts and the Future of Justice* (Oxford University Press, 2019) 62–63.
39 Legg and Bell (n 8) 138–9.

The ODR system may be linked with the court/tribunal system in a particular jurisdiction so that there is a seamless progression into that system as Figure 7 shows.[40]

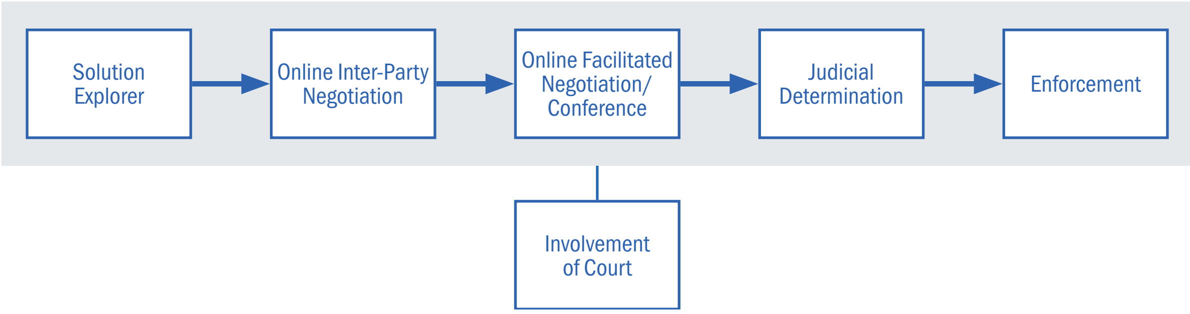


Figure 7: Integrated online court resolution of disputes

Alternatively, the system may generate forms and other documents (such as letters of demand, pleadings or affidavits) that the user can employ to commence legal proceedings. The aim of the expert system in step 1 and the ADR in step 2 is to resolve as many disputes as possible, as early as possible. The more self-resolution that occurs, the quicker and cheaper the process is for both the user and provider.

One of the most prominent and acclaimed examples of successful ODR is the British Columbia's Civil Resolution Tribunal.[41] In 2012, the British Columbia government passed the Civil Resolution Tribunal Act with the goal of using technology and ADR to increase access to justice for British Columbians with small claims and condominium property disputes. The CRT started with strata property disputes, expanded to small claims under $5,000 and then to motor vehicle accident and injury claims below $50,000. The small claims jurisdiction is planned to be gradually increased to claims under $25,000.[42] The CRT is comprised of 28 tribunal members supported by a staff of 78 employees.[43]

The CRT is composed of four steps as set out in Figure 8: The Civil Resolution Tribunal Process.

Figure 8: The Civil Resolution Tribunal Process

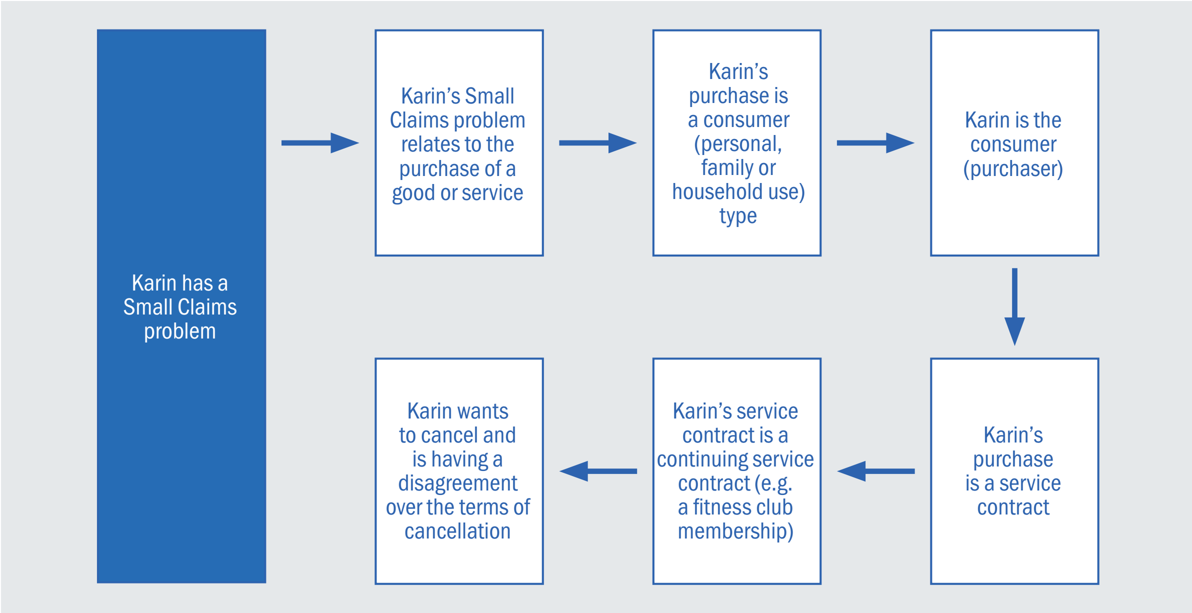
40 Figure taken from Felicity Bell et al, 'The Use of Technology (and other measures) to Increase Court Capacity: A View from Australia' (2021) 8(2) *International Journal of Online Dispute Resolution* 102, 107.

41 For other examples see Legg and Bell (n 8) 139-49.

42 Shannon Salter and Darin Thompson, 'Public-Centred Civil Justice Redesign: A Case Study of the British Columbia Civil Resolution Tribunal' (2016) 3 *McGill Journal of Dispute Resolution* 113; Shannon Salter, 'Online Dispute Resolution and Justice System Integration: British Columbia's Civil Resolution Tribunal' (2017) 34(1) *Windsor Yearbook of Access to Justice* 112.

43 Civil Resolution Tribunal, *2020/2021 Annual Report* 39-40.

The first step employs an expert system called the Solution Explorer. Solution Explorer uses interactive questions and answers to give people tailored legal information as well as tools and resources to assist them in answering the questions asked. It also classifies the dispute and provides the appropriate online application form. For example, the expert system may ask questions to diagnose a person's problem by narrowing it from a wide domain down to a more granular level as follows:



The narrowing of the issue also enables the expert system to deliver targeted information to the user about the problem or issue, including the identification and explanation of potentially relevant rights and obligations. The second and third steps involve the plaintiff and defendant being filtered through a structured negotiation session and 'facilitation' aided by a case manager for coming to an agreement. Failing the agreement, parties may apply to have a CRT Member adjudicate the matter. The adjudication does not require an in-person hearing as communications technology facilitates the hearing. The Solution Explorer was used 160,527 times from 13 July 2016 to 31 March 2021. In 2020/2021 the average time to resolution for all dispute types was 85.8 days and the median time to resolution was 59 days for all dispute types.[45]

Another example is the English Traffic Penalty Tribunal (TPT). The TPT decides motorists' appeals against Penalty Charge Notices (PCNs), issued by local authorities and charging authorities in England (outside London) and Wales, for parking and traffic contraventions. The Tribunal comprises 30 part-time adjudicators who are judicial officers working remotely with the support of 14 administrative staff. The process employs 'Triage questioning' for appellants during the appeal registration process which guides them through the information they need to provide to initiate an appeal, including about themselves, the vehicle and the PCN. Other technology is also used. The process provides for the upload of evidence, such as photographs and videos, to PDFs of documents, to screen captures of WhatsApp messages. Appellants have the option to select either:

(1) an e-decision: A TPT Adjudicator will decide the appeal without a hearing or talking to the parties, often asking questions in a message and the parties replying promptly.

(2) a telephone hearing: the motorist can ask for teleconference with the adjudicator and an Authority representative usually taking part.[46]

In the United Kingdom, an online portal known as Money Claim Online ('MCOL') has, since 2002, facilitated simple, small claims of £100,000 or less without the need to enter a court building or engage a solicitor.[47] A comprehensive practice note, which supplements the *Civil Procedure Rules*, *Practice Direction 7E – Money Claim Online*,[48] delineates the rules and procedure applicable to the MCOL, including the types of claims that can be made (dir 4) and the way that a claim ought to be commenced (dir 5).

44 Salter and Thompson (n 42) 129.
45 Civil Resolution Tribunal, *2020/2021 Annual Report* 12, 28,
46 Traffic Penalty Tribunal, *Revolutionising a service* (November 2019); *Traffic Penalty Tribunal* (Web Page) <https://www.trafficpenaltytribunal.gov.uk/>.
47 HM Courts & Tribunals Services, 'Money Claim Online (MCOL) – User Guide for Claimants' <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/762843/mcol-userguide-eng.pdf>.
48 Practice Direction 7E – Money Claim Online, 1 October 2020 <https://www.justice.gov.uk/courts/procedure-rules/civil/rules/part07/pd_part07e>.

A separate portal, made public in 2018 and known as the Civil Money Claims portal,[49] allows applicants to make a claim if the value of their loss is less than £10,000.[50] A 2017 survey of the Civil Money Claims portal trial participants reported that 80% of users found the service easy to use.[51]

The programs take users through the eligibility requirements necessary to make a claim before determining whether their matter is suitable for the MCOL or the Civil Money Claims portal. If the case is defended and certain automatically generated documents are filed via the system, the claim may go to mediation or the local court. However, non-response or a willingness by the defendant to pay the sum can facilitate a 'judgment' through the money claim online portal. The user inputs the terms of the 'judgment' (e.g. the method of payment, whether it is to be paid by instalments) to be confirmed by the court. The portal can be used to issue a warrant in the event of non-payment.

3.3 Prediction of Litigation Outcomes

Given sufficient volumes of case law, it is possible to create machine learning models which can 'predict' the outcome of legal cases. There is quite a long history of statistical and computational modelling of legal cases,[52] and many early systems aimed to predict case outcomes through traditional statistic approaches that identified correlations between case features and case outcomes. Machine learning enables the identification of more complex relationships and patterns, although it may be more difficult to provide explanations for predictions (see section 2.10 Explainable Artificial Intelligence).

Outcomes of litigation can be predicted through both expert systems and machine learning techniques, although the logic underlying the respective technology can be quite different. An expert system gives answers based on known rules; for example, using legal rules to determine who is liable in a vehicle accident and what damages are payable.[53] Machine learning relies on patterns in historic data. As mentioned in section 2.6 Machine Learning, a simple machine learning decision tree model could predict the outcome of US Supreme Court decisions.[54] The data input into machine learning models like this could be comprised of many different features of a case. Features might be factors that would be known prior to the case being argued, such as whether a party is self-represented, whether a party is a corporation, the identity of the lawyers, the identity of the judge, and so on. Alternatively, the features could be information about the events which have given rise to the claim, such as the factual circumstances of the case. As it is time-consuming to identify and label features like this,[55] there is interest in machine learning programs which can themselves identify relevant features from a corpus of documents, weight them, and use this information to make predictions about new cases.[56]

Several research groups have built machine learning programs which have been able to predict the outcomes of decisions in various courts including the Australian Federal Court, the French Court of Cassation and the European Court of Human Rights (ECtHR). Similar machine learning programs have been developed in respect of discrete legal issues such as the outcome of securities fraud class actions[57] and intellectual property lawsuits.[58]

AI systems can, it seems, achieve very good accuracy. Sulea et al predicted case rulings at the French Supreme Court with an accuracy of 92%.[59] Katz et al, who targeted their analysis at predicting the outcome of US Supreme Court decisions,[60] were able to predict specific judicial votes in 240,000 instances with a 71.9% accuracy.[61] However, these results are not always as useful as they may seem. In relation to the studies which focussed on the EctHR, one flaw in the design was that the 'facts' used in training the model were the judge's summary of the facts in the judgments themselves.[62] If facts are selected or presented in a manner that favours one outcome, the system might learn to associate such signalling with an outcome rather than the 'raw facts' or facts presented by the

49 Lucy Frazer MP, Parliamentary Under-Secretary of State for Justice to Bob Neill MP, Chair of the Justice Select Committee, House of Commons, 'HM Courts & Tribunals Service New Online Civil Claims Pilot Rolled Out' (April 2018) <https://www.parliament.uk/documents/commons-committees/Justice/correspondence/Lucy-Frazer-HMCTS-online-civil-claims-pilot.pdf>.
50 Practice Direction 51R - Online Civil Money Claims, 8 December 2021; 'Make a Money Claim Online' (*GOV.UK*) <https://www.gov.uk/make-money-claim>.
51 Frazer (n 49).
52 See Ibid. 108.
53 This seems to be the basis of the Singapore system. See 'Motor Accident Claims Online | Motor Accident Claims Online (MACO) Helps Parties Involved in Motor Accidents Make Informed Decisions about Their Motor Accident Claims.' <https://motoraccidents.lawnet.sg/About>.
54 Ruger et al (n 14).
55 In the United States, a comprehensive database of Supreme Court decisions with labelled characteristics is maintained. See Harold J Spaeth and James L Gibson, 'United States Supreme Court Judicial Database Terms Series' (4 November 2005) <https://www.icpsr.umich.edu/web/ICPSR/series/86>. No equivalent database exists in Australia.
56 Ashley (n 52).
57 Blakely B McShane et al, 'Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits' (2012) 9(3) *Journal of Empirical Legal Studies* 482.
58 Mihai Surdeanu et al, 'Risk Analysis for Intellectual Property Litigation' in *International Conference on Artificial Intelligence and Law* (2011) 116, 116–120.
59 Octavia-Maria Sulea et al, 'Predicting the Law Area and Decisions of French Supreme Court Cases' [2017] *arXiv:1708.01681 [cs]* <http://arxiv.org/abs/1708.01681>.
60 Daniel Martin Katz, Michael J Bommarito Ii and Josh Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' (2017) 12(4) *PLoS ONE*.
61 Ibid.
62 Nikolaos Aletras et al, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 PeerJ Computer Science 93; Masha Medvedeva, Michel Vols and Martijn Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights' (2020) 28 Artificial Intelligence and Law 237, 256.

parties. This is an inherent limitation in existing AI prediction methods; they may succeed in predicting an outcome given certain defined variables, but they have not been demonstrated to be effective at predicting new cases.[63] Further, accuracy ratings need to be assessed against a default of 50% (coin toss) or higher (if 80% of matters on a particular topic are dismissed, then one can achieve 80% accuracy by guessing that every matter will be dismissed). In this context, a 79% accuracy rating (as achieved by Aletras et al in relation to EctHR decisions) may not be as successful as it seems. Pasquale and Cashwell make the point that the study should not be used for purposes such as decision-making or even triage of matters because it necessarily takes irrelevant factors into account.[64]

Compared to legally relevant features such as case facts, it is technically more straightforward to extract information pertaining to characteristics such as the identity of the judge and lawyers involved in a case and use this to predict judicial outcomes. One study found that, using only the names of the judges on the bench, the outcome of EctHR cases could be accurately predicted 65% of the time.[65] This number was even higher in relation to some types of cases; European Convention on Human Rights art 13 (right to an effective remedy) outcomes could be predicted 79% of the time. Such tools could be used to measure statistical differences between judges – for example by ranking judges according to whether they are statistically more likely than other judges to find for plaintiff/applicant or defendant/respondent in a particular category of cases. This can also identify 'outlier' judges who may never or rarely make particular findings. Courts in China use AI to give judges a warning if their judgment does not match what a database has predicted it ought to be.[66] An outlier judgment warning has been added in some courts, generating a warning to a judge's superiors, mostly in the context of criminal sentencing.[67]

However, as discussed in section 4.3 Impartiality and Equality Before the Law, care must be taken before treating such statistics as evidence of bias. In Australia, this will not be sufficient to demonstrate apprehended bias for the purpose of a disqualification application. Nevertheless, this type of analysis is being done; for example, researchers at Macquarie University have analysed migration decisions, though emphasising that their goals are descriptive rather than necessarily predictive.[68]

Summary

Using information about cases or even the text of decisions themselves as inputs, machine learning programs can predict case outcomes, some with good accuracy. However, there are limitations, and concerns about how such predictions might impact the delivery of justice.

A variety of products claim to be able to predict the outcome of civil litigation (win/lose), damages awards and costs using variables such as the identity of the judge and the nature of the case. Such tools can assist with litigation strategising and costing, understanding an opponent's common strategies, and making decisions about settlement.[69] There are also tools used for specific purposes, such as litigation funding.[70]

The use of statistics and prediction is also relevant in the context of criminal proceedings, and in particular to predict sentencing outcomes.[71] These work on a different principle to risk assessment tools used in sentencing (see section 3.4 Criminal Sentencing and Risk Assessment Tools), being based on commonalities with historic precedents.

When interrogating the usefulness and accuracy of any tool, it is important to understand how it works (including how data is sourced and whether it represents the phenomenon being studied) as well as how success is measured, and in particular, whether an independent evaluation has been conducted. Machine learning can sometimes identify patterns that are an artifact of the data or methodology used rather than genuinely predictive. It is also crucial to recognise the distinction between the use of probabilistic predictions of outcomes for academic study or for parties' information and their use in replacing some of the decisions they aim to predict.

63 John Morison and Adam Harkens, 'Re-Engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making' (2019) 39(4) *Legal Studies* 618, 623.
64 Frank Pasquale and Glyn Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' (2018) 68(1) *University of Toronto Law Journal* 63.
65 Masha Medvedeva, Michel Vols and Martijn Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights' (2020) 28 *Artificial Intelligence and Law* 237, 256, 259-262.
66 Brian M Barry, *How Judges Judge: Empirical Insights into Judicial Decision-Making* (Informa Law from Routledge, 2021) 278.
67 Meng Yu and Guodong Du, 'Why Are Chinese Courts Turning to AI?' (19 January 2019) *The Diplomat* <https://thediplomat.com/2019/01/why-are-chinese-courts-turning-to-ai/>; Sourdin (n 11) 14.
68 'Study Raises Questions about Visa Appeals', *AMES Australia* (12 March 2020) <https://amesnews.com.au/latest-articles/study-raises-questions-about-visa-appeals/>; Hagar Cohen, 'Asylum Seekers Have Slim Chance with One Particular Judge, New Figures Reveal', *ABC News* (online, 6 September 2019) <https://www.abc.net.au/news/2019-09-06/almost-99-per-cent-fail-when-heard-by-judge/11457114>.
69 Marketing, 'Legal Analytics - Quickly Uncover Strategic Information', *Lex Machina* <https://lexmachina.com/legal-analytics/>.
70 'Legalist' <https://www.legalist.com/philosophy>.
71 Duncan I Simester and Roderick J Brodie, 'Forecasting Criminal Sentencing Decisions' (1993) 9 *International Journal of Forecasting* 49.

3.4 Criminal Sentencing and Risk Assessment Tools

Some US jurisdictions[72] use AI systems to augment and, in part, replace judicial discretion in the prediction of the likelihood that an accused (re)offends in the context of criminal bail and sentencing decisions.[73] For example, the Correctional Offender Management Profiling for Alternative Sanctions tool (COMPAS)[74] is used to conduct a risk assessment by drawing on the historical data of offenders and analysing that data to produce an output based on the particular offender's conduct and background. COMPAS integrates 137 responses to a questionnaire, which includes questions ranging from the clearly relevant consideration, 'how many times has this person been arrested before as an adult or juvenile', to the more opaque 'do you feel discouraged at times'.[75] Importantly, the code and processes underlying COMPAS is secret, and so not known to the prosecution, defence or judge.

COMPAS was developed in 1998, and can be used to predict, first, the likelihood that an accused fails to appear for trial (the 'Pretrial Release Risk' scale), second, the likelihood that an offender commits subsequent offences (the 'General Recidivism' scale), and third, the likelihood that an offender commits a violent act in the future (the 'Violent Recidivism' scale).[76] The outcome of each assessment can be used by a court to determine, for example, whether the accused should be released on bail pending trial or be subject to a suspended sentence (recognisance release order) in lieu of a custodial sentence. COMPAS, and risk assessment tools like it, predict the future behaviour of individuals who are either accused of criminal wrongdoing or are incarcerated having been convicted of a crime. Factors that risk assessment tools might take into account include education and employment, family, socioeconomic and geographical background, and association with convicted criminals by way of family or broader networks.

Supporters claim that COMPAS can determine whether an offender has a high likelihood of recidivism and that the program supports judicial decisions as to bail and sentencing on that basis. Many US jurisdictions allow, and some go so far as to require, judicial use of COMPAS or similar tools; since its development, COMPAS has been used to assess over one million offenders.[77] Indeed, the recently passed and incredulously named *Formerly Incarcerated Reenter Society Transformed Safely Transitioning Every Person Act* (the 'First Step Act') contains sections requiring the Attorney General to develop and release a risk and needs assessment system to determine the recidivism risk and violent or serious misconduct risk of each prisoner (being minimum, low, medium or high).[78]

There are several well-publicised instances of COMPAS impact on an accused and their liberty. In 2013, Paul Zilly was accused, tried and convicted in Wisconsin of stealing a lawnmower, among other tools, which he intended to sell for parts. The prosecution, together with Zilly's attorneys, agreed a plea deal which recommended one year in a county jail, and a subsequent supervision order. Presiding Judge James Babler stated at appeal that he would likely have sentenced Zilly to 18 months' incarceration. However, on the basis of COMPAS, which designated Zilly's likelihood of re-offending at 'about as bad as it could be', Judge Babler rejected the plea deal and sentenced Zilly to two years' imprisonment.[79]

COMPAS has faced a superior court challenge in the US. In 2013, Eric Loomis was charged and convicted in relation to a drive-by shooting. The Circuit Court noted that COMPAS had indicated that Loomis had a high risk in each of the pretrial recidivism, general recidivism and violent recidivism scales. On appeal, the Supreme Court of Wisconsin was asked whether the use of the COMPAS tool in sentencing violates a defendant's right to due process, either because the secret nature of COMPAS prevents defendants from challenging the assessment's scientific validity, or because COMPAS assessments take gender into account.[80] Justice Bradley, in delivering the reasons of the Court, held that the use of COMPAS by a court was permissible, so long as the judge made the final determination as to the sentence, and the judge is notified of the tool's limitations, namely that:[81]

72 For example, COMPAS has been used in at least the states of Florida, New York, Wisconsin and California. Keith Kirkpatrick, 'It's Not the Algorithm, It's the Data' (2017) 60(2) *Communications of the ACM* 21. Other jurisdictions use similar tools; see the California Static Risk Assessment and the Ohio Risk Assessment System. For a complete list, see 'AI in the Criminal Justice System – EPIC – Electronic Privacy Information Center' <https://epic.org/issues/ai/ai-in-the-criminal-justice-system/>.

73 Carolyn McKay, 'Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making' [2019] *Current Issues in Criminal Justice* 1.

74 Developed by Equivant (previously Northpointe): 'Home - Equivant', *Equivant* <https://www.equivant.com/>.

75 Julia Angwin, 'Sample COMPAS Risk Assessment - COMPAS "CORE"', *ProPublica* <https://www.propublica.org/documents/item/2702103-Sample-Risk-Assessment-COMPAS-CORE>.

76 Northpointe, *Practitioner's Guide to COMPAS Core* (19 March 2015) 27–28 <https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>.

77 Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4(1) *Science Advances* eaao5580.

78 *Formerly Incarcerated Reenter Society Transformed Safely Transitioning Every Person Act 2018* H. R. 5682, 3631–3633.

79 Alyssa M Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103 *Iowa Law Review* 303; Derek Thompson, 'Should We Be Afraid of AI in the Criminal-Justice System?' (*The Atlantic*, 20 June 2019) <https://www.theatlantic.com/ideas/archive/2019/06/should-we-be-afraid-of-ai-in-the-criminal-justice-system/592084/>; Julia Angwin and others, 'Machine Bias' [2016] *ProPublica* <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

80 *State of Wisconsin v Loomis* (2016) 881 N.W.2d 749 (Ann Walsh Bradley J) ('*Loomis*'). Cert denied, 137 S Ct 2290 (2017).

81 Ibid [66].

1. the method by which the risk scores were determined could not be disclosed to the court for proprietary reasons;

2. the tool compares a defendant to a national sample, and there was no evidence that this method was valid for a local (Wisconsin) population;

3. some studies have raised questions about whether the tool might give minorities a generally higher risk score; and

4. tools such as COMPAS should be 'constantly monitored and re-normed for accuracy' as population data changes.

The judge must consider defendant arguments countering the supposed risk he or she poses according to the COMPAS tool.[82] However, defence counsel has no correlated right to challenge the accuracy or methods of the COMPAS program,[83] neither of which are known to the defendant or judge.[84] In dealing with the ground of appeal related to COMPAS' use of gender as a factor, the Supreme Court of Wisconsin held that 'COMPAS's use of gender promotes accuracy that ultimately inures to the benefit of the justice system including defendants'.[85] Ultimately, the Court held that the tool could be used in proper circumstances, but cannot be used to determine whether an offender is incarcerated or to determine the severity of the sentence, or as the determinative factor in deciding whether an offender ought be released on a supervision order into the community.[86] In 2017 Loomis' petition to the US Supreme Court was denied.[87]

In a 2016 investigation, the nonprofit ProPublica looked at about ten thousand criminal defendants in Broward County, Florida, whose penalty consequent on the finding of criminal guilt had been, at least in part, informed by COMPAS. ProPublica's analysis found that African American defendants were at an increased risk of receiving a false positive COMPAS score (meaning that they were more likely to be flagged as high risk despite not *in fact* being high risk), whereas white defendants were more likely to receive a false negative COMPAS score (meaning that they were more likely to be flagged as low risk despite not *in fact* being low risk).[88]

While the finding that the rates of false positives and false negatives are correlated to racial characteristics does not, necessarily, reflect an *inherent* bias in the program/algorithms itself, it instead is a reflection of the *human* bias inherent in the data from which the program was trained. Notably, COMPAS' developers claim that race, as such, is not a factor that the model takes into account. In other words, a defendant is unlikely to have to identify their race for the purpose of the COMPAS questionnaire. Instead, the answers to other questions may serve as proxies for race – for example, where an offender's place of birth or residency contains a high proportion of people from a minority background who are over-policed and harshly sentenced due to human bias, the COMPAS system may indicate a higher risk score. If the program had, as training inputs and outputs, recognized a link between, say, the postcode a defendant lived in and the sentence they received, the program would also form that link as indicative of the process of reasoning it should undertake in calculating risk scores.

Looking beyond concerns about the COMPAS system itself, the usefulness of AI in sentencing will depend on how sentencing decisions are to be made. For example, a majority decision of the Australian High Court has noted (in the context of the use of guideline judgments for sentencing):

> *The production of bare statistics about sentences that have been passed tells the judge who is about to pass sentence on an offender very little that is useful if the sentencing judge is not also told why those sentences were fixed as they were.*[89]

---

82 Ibid [56].
83 McKay (n 73) 11.
84 *Loomis* (n 30) [51].
85 Ibid [86].
86 Ibid [98].
87 *Loomis v Wisconsin* , 26 June 2017, Docket no 16-6387.
88 Angwin et al (n 79); cf Matthew G Rowland, 'Technology's Influence on Federal Sentencing: Past, Present and Future' (2020) 26 *Washington and Lee Journal of Civil Rights and Social Justice* 565, 611 who argues that a 'single report or study alone is not enough to provide a definitive assessment of the technology'. See further 'Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing', UCLA Law Review (19 February 2019) <https://www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/> ('*Injustice Ex Machina*') who says that the inaccuracy of the false positive rate is a necessary trade-off for the accuracy of the true positive rate, and so ultimately comes down to a developer's notion of justice and fairness as a balance between defendant and community interests.
89 *Wong v The Queen* (2001) 207 CLR 584, [59].

An AI system based purely on quantitative considerations would be similarly useless in a system based on individualized sentencing. The benefit of the Judicial Information Research System (JIRS), the database of sentencing information maintained by the Judicial Commission of NSW,[90] is that it combines statistics about sentences by reference to the relevant offences but also includes details about the nature of the offence and the defendant.

Summary

Risk assessment tools which use data-driven inferencing have proliferated in the US criminal justice system. These tools are often proprietary meaning that their operation is opaque, and it is difficult to challenge their functioning. US courts have nevertheless held that these risk assessments may be used in judicial decision-making in some circumstances.

3.5 Automated Decision-Support and Decision-Making

AI systems can inform, augment or even entirely replace judicial discretion. Depending on the purpose of the system, and the safeguards thought necessary to be built into it, human oversight can range from human (or technology)-in-the-loop to full autonomy (see section 2.3 Automation). As explained above in Prediction of Litigation Outcomes, the techniques which can be used to make predictions about the outcome of litigation could also be used for triage or even to automate decision-making – but this raises a number of issues about due process and the rule of law.

In terms of practical impact, more work has been done on systems to support non-judicial decision-makers (such as administrative officials). While still in their early stages, the below examples chart various attempts to automate not just the processes of the judiciary, but the decision-making of judges themselves.

An Australian example of decision support is the Bail Assistant program being developed by the Judicial Commission of NSW which seeks to guide decision-makers through the complexities of the *Bail Act 2013* (NSW). There are plans to use the data from bail decisions to train a machine learning system which could then predict bail decisions.[91]

Another example in the judicial context is the EXPERTIUS system in Mexico,[92] which advises 'novice' judges and clerks as to whether a plaintiff is eligible for a pension in addition to the quantum of that pension.[93] The program takes users through three modules; first, giving them an opportunity to understand the process itself (the tutorial module); second, giving the user a space to provide evidence in support of their case in addition to assigning 'weights' to each piece of supporting documentation (the inferential module); and third, allowing the user to determine the amount of the pension they are entitled to given specified socio-economic criteria (the financial module).[94]

The UK has worked on creating an entirely online court which would handle some summary offences, allow offenders to enter a guilty plea, and produce a pre-determined penalty, all without the involvement of a magistrate.[95] Consequently, a matter could be dealt with in an entirely automated way with no human oversight whatever. The jurisdiction of the court would focus on strict liability summary offences that do not attract a penalty of imprisonment, such as fare evasion and possession of certain equipment without a licence.[96] The offender would be presented with the evidence put against them and the consequence of entering a guilty plea. They would have the capacity to contest the charge; however, some commentators, including the then chair of the Bar Council, Andrew Langdon QC, raised concerns that a defendant may choose to enter a plea of guilt 'out of convenience'.[97] An offender could elect for their matter to be heard by a human magistrate and, in that sense, the online procedure was entirely voluntary. By focussing on low-level offences, the proposed online court was thought to be capable of delivering a

90 See 'Judicial Information Research System (JIRS)', Judicial Commission of New South Wales, <https://www.judcom.nsw.gov.au/judicial-information-research-system-jirs/>.

91 The Hon TF Bathurst, 'Sir Maurice Byers Lecture 2021: Modern and Future Judging', 3 November 2021, [48].

92 E Cáceres, 'EXPERTIUS: A Mexican Judicial Decision-Support System in the FIeld of Family Law' in EBE Francesconi, G Sartor and D Tiscornia (eds), *Legal Knowledge and Information Systems* (IOS Press, 2008) 78, 87.

93 Davide Carneiro and others, 'Online Dispute Resolution: An Artificial Intelligence Perspective' (2014) 41 Artificial Intelligence Review 211, 227–228; E Cáceres, 'EXPERTIUS: A Mexican Judicial Decision-Support System in the FIeld of Family Law' in EBE Francesconi, G Sartor and D Tiscornia (eds), *Legal knowledge and information systems* (IOS Press 2008) 78.

94 Davide Carneiro and others, 'Online Dispute Resolution: An Artificial Intelligence Perspective' (2014) 41 *Artificial Intelligence Review* 211, 227–228.

95 Joshua Rozenberg QC, 'Automatic Online Conviction - The Online Court: Will IT Work?' (*The Legal Education Foundation*) <https://long-reads.thelegaleducationfoundation.org/automatic-online-conviction/>; Owen Bowcott, 'Government's £1bn Plan for Online Courts 'Challenges Open Justice'' The Guardian (15 March 2017) <http://www.theguardian.com/law/2017/mar/15/governments-1bn-plan-for-online-courts-challenges-open-justice>.

96 UK Ministry of Justice, *Transforming Our Justice System: Assisted Digital Strategy, Automatic Online Conviction and Statutory Standard Penalty, and Panel Composition in Tribunals* (Government Response No Cm 9391, 2017) 16; see also Jane C Donoghue, 'Reforming the Role of Magistrates: Implications for Summary Justice in England and Wales' (2014) 77(6) *The Modern Law Review* 928.

97 Owen Bowcott, 'Government's £1bn Plan for Online Courts 'Challenges Open Justice'' The Guardian (15 March 2017) <http://www.theguardian.com/law/2017/mar/15/governments-1bn-plan-for-online-courts-challenges-open-justice>.

service that would be 'just, proportionate, accessible to all and works better for everyone'.[98] The proposal stalled after the 2017 general election and does not seem to have been revived.[99]

The Estonian Ministry of Justice has sought to automate the adjudication of small contract disputes.[100] So-called 'AI judges' would be used to clear a backlog of cases with the intention of giving human judges more time and resources to deal with complex disputes. Reportedly, the project will 'adjudicate small claim disputes under €7,000. … In concept, the two parties will upload documents and other relevant information, and the AI will issue a decision that can be appealed to a human judge'. Ott Velsberg, Estonia's chief data officer, explained his confidence in the success of the automated system on Estonia's familiarity with virtual processes such as electronic voting and digital tax filing.[101] That this proposal has originated in Estonia is perhaps unsurprising, given prior suggestions by the nation's digital adviser that AI systems be granted some form of separate legal personality,[102] and the current use of deep-learning methods to determine farmer subsidies.[103] Estonia's announcement forms a part of the nation's broader agenda of civic digitisation, and the formation of online dispute resolution mechanisms processes.

3.6 Automated E-Filing

Electronic filing (e-filing) of documents in court/tribunal proceedings has become ubiquitous in modern court systems. Most standard e-filing systems use expert or rule-based systems (see section 2.2 Expert Systems and Traditional Programming) but AI may play a role in the future of e-filing.

E-filing is intended to reduce or eliminate reliance on physical documents to run a case. By April 2019, the UK Crown Court had reportedly saved over 100 million sheets of paper after moving to e-filing. Storing and locating documents is easier when they are in electronic format. Further, the capacity of parties, lawyers and judges to search lengthy documents for particular words or phrases has become near instantaneous through the use of searchable files, and the ability to navigate between relevant documents has been facilitated through the use of hyperlinked documents.[104]

E-filing may also decrease errors in the filed documents themselves and speed up court processes. The UK Crown Court reported that filing errors in divorce matters reduce from 40% to less than 1%, and the speed in which online civil money claims have been able to be issued has reduced from 15 days under the paper system to 10 minutes under the digital one.[105]

While e-filing increases efficiency in court administration, there may be no human administrator identifying errors. The benefit of e-filing is that the system should be able to identify whether a document has been correctly prepared and, if it has, accept that document for filing and carry out any consequent steps (for example, creating a sealed version of the document and automatically sending it to the parties to the dispute and the chambers of the judge).

98 UK Ministry of Justice, 'Transforming Our Justice System: Assisted Digital Strategy, Automatic Online Conviction and Statutory Standard Penalty, and Panel Composition in Tribunals' (2017) Government Response Cm 9391 16 para 21(b).

99 Kerry Underwood, 'What Will Happen to the Prisons and Courts Bill?', *Dispute Resolution blog* (4 May 2017) <http://disputeresolutionblog.practicallaw.com/what-will-happen-to-the-prisons-and-courts-bill/>.

100 Victor Tangermann, 'Estonia Is Building a "Robot Judge" to Help Clear Legal Backlog', *Futurism* (online, 25 March 2019) <https://futurism.com/the-byte/estonia-robot-judge>.

101 David Cowan, 'Estonia: A Robotically Transformative Nation', *Robotics Law Journal* (Online, 26 July 2019) <http://www.roboticslawjournal.com/global/estonia-a-robotically-transformative-nation-28728942>.

102 Bartosz Troczyński, 'Estonia Plans the Boldest AI Regulations', *newtech.law* (23 October 2017) <https://newtech.law/en/estonia-plans-the-boldest-ai-regulations/>; cf State Chancellery, Estonia, *Estonian Artificial Intelligence: Report of the Expert Group on Deployment* (May 2019) <https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne.pdf> which recommended that Estonia adopts the approach of the EU rather than develop its own set of principles in relation to AI.

103 Michèle Finck, 'Automated Decision-Making and Administrative Law' in Peter Cane et al (eds), *Oxford Handbook of Comparative Administrative Law* (Oxford University Press, 2020) 5–6.

104 HM Courts & Tribunals Service, Ministry of Justice and Lucy Frazer QC MP, 'Digital Court System Saves Enough Paper to Cover Central Park Twice', *GOV.UK* (18 April 2019) <https://www.gov.uk/government/news/digital-court-system-saves-enough-paper-to-cover-central-park-twice>.

105 'Digital Court System Saves Enough Paper to Cover Central Park Twice | Ministry of Justice', *WIRED GOV* (18 April 2019) <https://www.wired-gov.net/wg/news.nsf/articles/Digital+court+system+saves+enough+paper+to+cover+central+park+twice+19042019111500?open>.

THINGS TO CONSIDER

Examples of rule-based e-filing systems

- UK Courts & Tribunals use the CE-File system. Legal professionals or self-represented litigants can upload files and monitor the progress of cases, online.[106] From 2019, CE-File was made mandatory in the Business and Property Courts throughout the UK.[107]
- In the US, the NextGen CM/ECF system, deployed in conjunction with the Public Access to Court Electronic Records program (PACER), is used in all appellant, district, and bankruptcy courts.[108] The NextGen CM/ECF system is similar to CE-File and acts as a comprehensive case management system.[109] PACER gives the public access to over 1 billion documents filed in more than 200 federal courts.[110]
- In Australia, the National Court Framework, adopted by the Australian Federal Court in 2014,[111] streamlined and synchronised the operation of State registries and the operation of individual judges' dockets. E-filing is now offered by most Australian courts.[112]

In counties of Florida, California and Texas, courts use a machine learning tool, Intellidact AI, developed by Computing System Innovations (CSI), to filter e-filed documents.[113] CSI claims that Intellidact is able to classify and extract data from documents automatically using continuous supervised machine learning (see section 2.6 Machine Learning).[114] In 2020, the Florida county court reported that its goal was that at least 85% of all e-filed documents would be produced by Intellidact.[115] Intellidact uses machine learning to 'read' filed documents, extract relevant information, use that information to fill out docket sheets to be put into the case management system, and finally make those documents publicly available. Where a document does not fit into a category in the training data, the system puts it into a separate folder for human review. Ordinarily, however, the system operates without any human oversight,[116] and consequently e-filing is available continuously and not only when the court is open.[117] This system, as well as a similar system used in initiating criminal proceedings in Okaloosa County, Florida, can also automatically redact private or sensitive information before publishing the filed documents. CSI's CEO has said that Intellidact automatically processed 75–80% of all documents filed without human intervention.[118]

Summary

Automated e-filing systems may use rules-based systems or machine learning. The goal is to expedite filing as well as reduce or eliminate the use of paper documents.

---

106 'CE-File System Information and Support Advice', *GOV.UK* <https://www.gov.uk/guidance/ce-file-system-information-and-support-advice>.
107 'Electronic Filing in the Business and Property Courts Outside of London' <https://www.judiciary.uk/announcements/electronic-filing-in-the-business-and-property-courts-outside-of-london/>; 'CE-File System Information and Support Advice' (n 106).
108 'Electronic Filing (CM/ECF)', *United States Courts* (October 2021) <https://www.uscourts.gov/court-records/electronic-filing-cmecf>; 'NextGen CM/ECF | Central District of California | United States District Court' <https://www.cacd.uscourts.gov/e-filing/nextgen-cmecf>.
109 'NextGen CM/ECF | Central District of California | United States District Court' (n 108).
110 'Public Access to Court Electronic Records | PACER: Federal Court Records' <https://pacer.uscourts.gov/>.
111 Federal Court of Australia, 'Australian National Court Framework', *Federal Court of Australia* (12 November 2019) <https://www.fedcourt.gov.au/about/national-court-framework>.
112 This is variable – eg, in Tasmania, e-filing is achieved by sending an email attachment: James Allsop, 'Technology and the Future of the Courts' (Speech, TC Beirne School of Law, University of Queensland, 26 March 2019) 6.
113 Sean La Roque-Doherty, 'Artificial Intelligence Has Made Great Inroads, but Hasn't yet Increased Access to Civil Justice', *ABA Journal* (1 April 2021) <https://www.abajournal.com/magazine/article/artificial-intelligence-has-made-great-inroadsbut-not-as-far-as-increasing-access-to-civil-justice>; 'Charlotte County, Florida Uses AI To Process E-Filings – Smart Cities Connect' (13 May 2021) <https://smartcitiesconnect.org/charlotte-county-florida-uses-ai-to-process-e-filings/>. (Palm Beach and Charlotte County, Fl; Stanislaus County, Ca; Tarrant County, TX).
114 'Intellidact AI' <https://www.pdfa.org/product/intellidact-ai/>.
115 'Palm Beach Clerk Receives National Digital Innovation Award for Its Use of CSI Intellidact AI - Florida Court Clerks & Comptrollers' (4 January 2019) <https://www.flclerks.com/news/432529/Palm-Beach-Clerk-Receives-National-Digital-Innovation-Award-for-its-use-of-CSI-Intellidact-AI.htm>.
116 Gary Blankenship, 'Artificial Intelligence Is Already at Work in Some Clerks' Offices', *The Florida Bar* (1 August 2018) <https://www.floridabar.org/the-florida-bar-news/artificial-intelligence-is-already-at-work-in-some-clerks-offices/>.
117 'Charlotte County, Florida Uses AI To Process E-Filings – Smart Cities Connect' (n 113).
118 Sean La Roque-Doherty (n 47).

3.7 Triaging and Allocation of Matters

The use of e-filing has led naturally to virtual triaging and allocation processes. In many jurisdictions, triaging and allocations are done primarily or exclusively by court administrators or judges. Some systems keep ‘external’ material, such as documents put onto a court file and orders made in a proceeding, separate from ‘internal’ material, such as the mechanisms of the court in overseeing and determining a dispute. Other systems integrate all facets of a proceeding into a single online portal. For example, Israel’s Legal-Net, a ‘cloud-based comprehensive court administrative platform’,[119] centralises submissions of documents and motions, paying of court fees, planning of court calendars, official recording of witness details and appearances, facilitating the production of draft judgments, and tracking the progress of all matters before the courts.[120]

AI systems can be deployed in triaging or allocating matters within a court system in many ways. The Victor Project, created for use in Brazilian courts in 2018, seeks to reduce the substantial backlog being experienced by those courts.[121] In 2017 alone, 80.1 million matters were awaiting judicial determination in Brazil, many of which have been described as ‘routine and low value’.[122] The Federal Supreme Court of Brazil is using AI to increase the speed of case resolution, increase the precision and accuracy of matters, and facilitate appropriate allocation of human resources in the judicial system.[123] It does so by breaking down and classifying cases of so-called ‘generation repercussion’, being those of economic, political, social or legal relevance, into classes of cases which may be decided together. The program has reportedly reduced 40 minutes of judicial work into a program which takes 5 seconds to run.[124]

Similarly, AI systems can be used to direct the attention of the court.[125] Ryan Copus has shown how machine learning can be used to produce a ‘statistical precedent’, which asks ‘how frequently has the court reversed cases like this one?’.[126] Such systems could compare an individual outcome with general, or ‘standard’, jurisprudence, determine whether a case is ‘easy’ and requires limited attention, or ‘hard’ and so useful in developing the law. Statistical precedents could also flag decisions which depart widely from the norm, label certain unpublished opinions as ‘high-risk’, include cases relevant to an unstable statistical precedent in reporting publications, and assign more simple cases to lower-level court staff. Although this ‘statistical precedent’ is purely theoretical, it is easy to see how such processes may be useful. For example it could dispose with simple applications for leave to appeal a decision and instead an appellant could opt-in to the use of an AI system to determine whether there is a reasonably arguable case that the primary judge erred and so leave to appeal should be granted, such that if leave to appeal were denied the appellant would be faced with significantly decreased costs. Or it could help law reporters determine which cases ought to be published by determining which cases relate to areas of law with an ‘unstable’ statistical precedent or vary widely from the statistical precedent.

It is, of course, essential that AI systems used in case management comply with relevant procedural rules and are updated as those rules change. A failure to do this can lead to difficulties, as evident in *Hemmett v Market Direct Group Pty Ltd [No 2]* decided by the Supreme Court of Western Australia.[127] In that case, the software used did not provide for an actual Inactive Cases List as required to implement the scheme in *Magistrates Court (Civil Proceedings) Rules 2005* (WA) Pt 16A; thus the consequences prescribed in those rules for inactive cases did not apply.[128] Identifying and implementing procedural requirements, as they evolve, is thus essential.

Summary

Following on from the use of e-filing, triage and allocation of court matters could be automated. This might include classifying or directing cases by using patterns in existing court data.

119 Amnon Reichman, Yair Sagy and Shlomi Balaban, ‘From a Panacea to a Panopticon: The Use and Misuse of Technology in the Regulation of Judges’ (2020) 71 *Hastings Law Journal* 589, 597.
120 Ibid 598–601.
121 ‘Inteligência Artificial: Trabalho Judicial de 40 Minutos Pode Ser Feito Em 5 Segundos’, *Supremo Tribunal Federal* <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=393522>.
122 Susskind (n 38) 290.
123 Pedro HG Inazawa et al, ‘Project Victor’ (2019) 39(1) *Revista Computacao Brasil, Sociedade Brasileira de Computacao* 19, 20.
124 Daniel Willian Granado, ‘Artificial Intelligence Applied To The Legal Proceedings: The Brazilian Experience’ (2019) 5 *Revue Internationale de droit des données et du numérique* 103.
125 Ryan W Copus, ‘Statistical Precedent: Allocating Judicial Attention’ (2020) 73 *Vanderbilt Law Review* 605.
126 Ibid 611.
127 [2018] WASC 310.
128 Ibid.

3.8 Natural Language Processing

Natural language processing, often using machine learning, can recognise, process and analyse languages, and convert them into another form, such as audio to text. Briefly, 'since language is contextual, statistics are used to work out the probability of words appearing near one another in a text'.[129]

In Australia, services such as Auscript, Transcription Australia and Epiq provide courts with transcription services, some of which boast real-time transcription. Voice recognition and transcription can be automated and globally, the speech recognition market is expected to be worth at least USD18 billion by 2023. IBM has achieved a 5.5% word error rate (compared to the standard human error rate of 5.1%),[130] with a 'dramatic improvement in accuracy' driving the likelihood that court reporting will increasingly be an automated process.[131] VIQ Solutions reportedly uses AI transcription and in 2020 announced it had secured a contract for transcription services with Queensland's Department of Justice and Attorney-General.[132]

Some Chinese courts use real-time voice recognition to produce court transcripts.[133] iFLYTEK is a technology company used during some trials which translate real-time audio into Mandarin and English text.[134] In Shanghai, at least ten courts are piloting the complete replacement of judicial clerks with AI assistants, whose role it is to transcribe cases, pull files and present digital evidence.[135]

Summary

Natural language processing typically uses machine learning to analyse text. Its main use for courts is in voice recognition and transcription of court proceedings.

3.9 AI-Supported Legal Research

AI can assist people to find what they are looking for in amongst a trove of digital documents. The benefits of natural language processing over traditional keyword searching are illustrated by Google's ubiquity. Lay people seeking help to resolve a legal problem on their own generally start with a Google search.[136] Many lawyers likely take the same approach.

Some legal research providers market themselves with an AI focus, and most are using at least some AI techniques. Ross Intelligence described itself as building 'AI-driven products to augment lawyers' cognitive abilities',[137] including natural language searching and flags for 'bad law'.[138] However, most legal research tools use automation and/or machine learning to help researchers find and link to precedents related to a passage a case (or paragraph) they are reading and many also rely on natural language processing for queries. LexisNexis, for example, incorporates 'AI-powered features' in their legal research platforms.[139] Austlii also uses automation in NoteUp function, which finds documents relevant to the document being viewed.

Expert systems are also used in legal research. AustLII's Datalex platform enables statutes to be written in a machine-consumable format, allowing users to find out how a statute applies in a particular situation by answering a series of questions.[140] This example is also discussed below (see section 3.10 Rules as Code – Implications for the Judiciary). The benefit of this approach is that it outputs the reasons why a particular provision does or does not apply (with statutory references) which is likely faster than reading a statute from beginning to end.

---

129 Legg and Bell (n 9) 35.

130 'What Is Speech Recognition?', *IBM* (2 September 2020) <https://www.ibm.com/cloud/learn/speech-recognition>; George Saon, 'Reaching New Records in Speech Recognition', *Watson Blog* (7 March 2017) <https://www.ibm.com/blogs/watson/2017/03/reaching-new-records-in-speech-recognition/>.

131 President of Voice Information Associates Walt Tetschner, quoted in David Ward, 'The Latest on Speech Recognition', *The Journal of Court Reporting* (29 November 2016) <https://www.thejcr.com/2016/11/29/the-latest-on-speech-recognition/>.

132 'VIQ Solutions Announces Details of Queensland Department of Justice and Attorney General Six Year Contract', *VIQ Solutions* (16 December 2020) <https://viqsolutions.com/media-center/viq-details-queensland-six-year-contract/>.

133 Damian Taylor and Natalie Osafo, 'Artificial Intelligence in the Courtroom', *Law Gazette* (9 April 2018) <https://www.lawgazette.co.uk/practice-points/artificial-intelligence-in-the-courtroom-/5065545.article>.

134 Lin Haibin, Guodong Du and Meng Yu, 'Big Data, AI and China's Justice: Here's What's Happening', *China Justice Observer* (1 December 2019) <https://www.chinajusticeobserver.com/a/big-data-ai-and-chinas-justice-heres-whats-happening>.

135 Deputy President of Xuhui District People's Court Xu Shiliang, quoted in Sarah Dai, 'Shanghai Judicial Courts Start to Replace Clerks with AI Assistants', *South China Morning Post* (1 April 2020) <https://www.scmp.com/tech/innovation/article/3077979/shanghai-judicial-courts-start-replace-clerks-ai-assistants>.

136 'Seeking Legal Help Online: Understanding the "Missing Majority" Report', *Justice Connect* (November 2020) <https://justiceconnect.org.au/about/digital-innovation/missing-majority-report/>.

137 'About Us', *ROSS Intelligence* <https://rossintelligence.com/about.html>.

138 'Features', *ROSS Intelligence* <https://rossintelligence.com/features.html>.

139 'The Power of Artificial Intelligence in Legal Research' (October 2020) <https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/the-power-of-artificial-intelligence-in-legal-research, https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/the-power-of-artificial-intelligence-in-legal-research>.

140 'DataLex' (n 19).

3.10 Rules as Code – Implications for the Judiciary

RaC (see section 2.4 Rules as Code (RaC)) has been trialled in different jurisdictions, including New Zealand, New South Wales and France, but is not currently being used at scale in any jurisdiction. Although RaC is at a preliminary stage, the implications for courts and tribunals remain largely hypothetical but will be discussed in this section.

One RaC project was the creation of a machine-consumable version of the *Community Gaming Regulation 2020* (NSW). The computer code for this project is stored on a website called GitHub, where it can be publicly accessed (although this may not be easily interpretable by non-experts). Once rules are written in a machine-consumable format, the government or third parties can create applications that allow users to query the rules to understand how they apply to their own situation. For example, NSW Fair Trading has created such an application.[141] Another NSW RaC initiative is the Energy Saving Certificate (ESC) calculator which helps NSW building owners to determine their eligibility to participate in the *NSW Energy Savings Scheme (ESS)* under the *NABERS baseline method*.[142]

Beyond government, AustLII has developed ElectKB which is essentially a machine-consumable version of section 44 of the Australian *Constitution*.[143] The program is designed as a chatbot to assess whether an individual is eligible to stand as a member of Australia's Federal Parliament.[144]

Other jurisdictions are also developing or have developed RaC projects. The French government initiated the OpenFisca project which focuses on the domain of tax and social benefits.[145] The OpenFisca platform, also used in the NSW project on the *Community Gaming Regulation 2020*, cultivates the development of projects that codify rules and perform simulations of future public policy changes. It operates through an open-source access format which allows the general public to contribute to the development of the code, thus fostering transparency and access to the law.[146]

Med Aides is a social benefit simulator built using the OpenFisca platform which aims to inform French citizens on their eligibility to national and local social benefits.[147] Mon Entreprise is another RaC initiative built by the French government that offers a range of simulators designed to help business owners understand and comply with the rules associated with running a business in France.[148]

The New Zealand Better Rules Project has used RaC concepts in a variety of contexts to provide more efficient services. The Wellington City Council has developed a project looking at how to incorporate the Better Rules methodologies to implemented in the context of urban planning to help inform the new district plan.[149]

The transparency of the underlying rule-set in RaC projects means that any organisation can build its own interface through which users can query the rules.[150] Where machine-consumable versions of rules are used directly in government decision-making, transparency is useful to individuals considering challenging the decisions in administrative law.

There are longer-term implications of RaC for judges. For example, in future, a legislature may seek to give a machine-consumable version of legislation a formal status alongside the natural language (e.g. English) counterpart.[151] The advantage would be a reduction in compliance risks for organisations seeking to follow the rule, as well as (potentially) higher rates of compliance.[152] From a judicial perspective, this raises questions of what it might mean to 'interpret' instructions written for a computer. The rules of statutory interpretation assume that what is being interpreted is natural language text addressed to people. There are no equivalent rules for judicial interpretation of computer code that causes a machine to perform a series of steps. The judiciary therefore will have an important future role in how machine-consumable versions of rules are recognised and interpreted in the context of disputes.[153]

---

141 The application is available at: https://www.fairtrading.nsw.gov.au/community-gaming/community-gaming-regulation-check/_nocache.
142 The calculator is available at: https://www.nabers.gov.au/ratings/calculation-tools/nsw-esc-estimator.
143 James Mohun and Alex Roberts, *Cracking the Code: Rulemaking for Humans and Machines* (No 42, OECD, 2020) 12 <https://www.oecd-ilibrary.org/content/paper/3afe6ba5-en>.
144 The chatbot is available at: http://datalex.org/app/consultation?rulebase=http%3A%2F%2Faustlii.community%2Ffoswiki%2FDataLex%2FElectKB.
145 Mohun and Roberts (n 126).
146 Ibid; The English language version of the platform can be accessed at https://openfisca.org/en/.
147 'OpenFisca: Mes Aides' <https://fr.openfisca.org/showcase/mesaides/>. The social benefits simulator is available at: https://www.mesdroitssociaux.gouv.fr/votre-simulateur/accueil.
148 Mohun and Roberts (n 126); One of the simulators is available at: https://www.maboussoleaidants.fr/mes-aides-financieres. Note the links in this paragraph will take you to French websites. You can translate the pages to English simply by a left click on the page on a Chrome browser and then select translate.
149 Ibid; Hamish Fraser, 'What Is Better Rules?', *New Zealand Digital government* (20 December 2019) <https://www.digital.govt.nz/blog/what-is-better-rules/>.
150 Mohun and Roberts (n 126).
151 Ibid.
152 Lyria Bennett Moses, Janina Boughey and Lisa Burton Crawford, 'Laws for Machines and Machine-Made Laws' in Janina Boughey and Katie Miller (eds), *The Automated State: Implications, Challenges and Opportunites for Public Law* (Federation Press, 2021).
153 Mohun and Roberts (n 126).

4. The Impact of AI Tools on Core Judicial Values

This section looks at how the AI technologies described in section 3 have the capacity to both undermine and strengthen judicial values. Clearly, these values are wide and subject to differing interpretations and emphases. Without wishing to engage in debate about the nature of core judicial values and what these may encompass, this guide focuses on open justice, accountability, impartiality and equality before the law, procedural fairness, access to justice and efficiency. These values often overlap and interact with one another, and the context of AI tools is no exception. Yet, they are useful guiding points for understanding how AI technologies can impact the courts and judiciary in the Asia-Pacific and other parts of the world.

4.1 Open Justice

Open justice subjects court proceedings to public and professional scrutiny and is critical to public confidence in the judicial system.[154] Many AI tools can enhance open justice beyond what would have been possible in a traditional courtroom. A basic, but nonetheless important, example is where an instant and automated transcription or translation service enables parties or members of the public who do not speak the language used in a courtroom to understand the proceeding by way of the translation. Some scholars claim that automated decision-making systems, if 'correctly' designed, could reveal each step necessary to reach a judicial decision, providing more information about how a decision is reached than a traditional judgment.[155]

However, AI tools can also undermine open justice, enabled by public and professional scrutiny. Many of the technologies described in section 3 fail to provide detail as to their operation, either to the public, parties to litigation or even the judge presiding over a matter. Even e-filing raises this concern – an attempt to file a document may fail for reasons not related to published rules (such as where a file is too large) or for reasons that are obscure to both the litigant and the registry. While the implementation of modern technological tools in a judicial setting may seem justified in circumstances where the judge has the capacity to review or override automatically generated outcomes, that safeguard is substantially undermined where a judge cannot view or understand the reasoning of an AI system.

There are three significant obstacles to ensuring open justice with AI tools. First, those responsible for AI systems may decide not to share information about how they work, for reasons of operational secrecy, to protect commercial information or to protect the privacy of personal information in training data. For example, the owners and developers of the COMPAS risk-assessment and sentencing tool have declined to disclose the core methods and datasets used. This lack of transparency was the focus of Justice Abrahamson's concurring judgment in *Loomis*, where her Honour understood the 'court's lack of understanding' of the tool as a 'significant problem'.[156] Her Honour further observed that 'making a record, including a record explaining consideration of the evidence-based tools and limitations and strengths thereof, is part of the long-standing, basic requirement that a circuit court explain its exercise of discretion at sentencing'.[157] Without the tool's mechanisms being public, the population against whom COMPAS could be instrumentalised lack 'a transparent and comprehensible explanation of the sentencing court's decision'.[158] Intentional secrecy produces significant harm to open justice (and accountability, as we discuss below),[159] and 'undermine[s] trust in AI and algorithmic outputs'.[160]

The second obstacle to ensuring open justice is that not everyone can understand artefacts that might explain the operation of an AI system. For example, source code for a computer program may not be understandable by those untrained. The provision of reasons is a judicial imperative which lies at the core of the communication between the courts and parties whose legal interests are affected.[161] The tension between language and technology is well illustrated by hypothetical publication of COMPAS code. No reasons for conclusions reached would be evident, and very few members of the general public would be able to achieve even basic comprehension of its meaning.[162]

154 *Hogan v Hinch* [2011] HCA 4; 243 CLR 506 [20].
155 Susskind was describing first wave artificial intelligence systems: Richard E Susskind, *Expert Systems in Law: A Jurisprudential Inquiry* (Clarendon Press 1987) 114–115; See further Susskind (n 38) 288.
156 *Loomis* (n 80) 774.
157 Ibid 133, 141.
158 Ibid 142.
159 Katherine Freeman, 'Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis' (2016) 18(5) *North Carolina Journal of Law & Technology* 75; Finck (n 103) 9.
160 Rowland (n 88) 611.
161 See, eg, *Wainohu v New South Wales* (2011) 243 CLR 181; Judge Henry J. Friendly, United States Court of Appeal for the Second Circuit, 'Some Kind of Hearing' (1975) 123 *University of Pennsylvania Law Review* 1267, 1291–1292; cf *Oakley v South Cambridgeshire District Council & Anor* [2016] EWHC 570 (Admin) which held that there is no general common law duty to give reasons in the UK [30]. However, the Court accepted that a duty to give reasons could arise in some circumstances [41], presumably alluding to situations in which procedural fairness would require it.
162 Michèle Finck, 'Automated Decision-Making and Administrative Law' in Peter Cane and others (eds), *Oxford Handbook of Comparative Administrative Law* (Oxford University

Most lawyers would not be able to interpret it. Ignorance of the process by which a legal decision has been made can be disempowering and make litigants vulnerable,[163] and constitute a denial of natural justice.[164]

Third, open justice can be undermined, because some systems are so complex that a process-based explanation is unhelpful in understanding a system's outputs. For example, an explanation of connections in an artificial neural network is as unhelpful in understanding the system as is a neuron-by-neuron description of a human brain in understanding the reasons for a complex (or even simple) decision made by a human. Even with the data science training to understand the process (overcoming the second obstacle above), the human mind cannot clearly 'see' a complex neural network with a hundred thousand layers (see section 2.10 Explainable Artificial Intelligence). The only way to understand the system here is to treat it as a 'black box' – to look at what goes in and what comes out and then to draw conclusions about its behaviour (see section 2.9 Technological 'black box'). For example, one could evaluate whether a system makes similar recommendations for men and women by inputting data on a random population sample comprising these genders.

The risk that lack of AI explainability poses to the judicial value of open justice is highlighted by entirely automated dispute resolution (see section 3.5 Automated Decision-Support and Decision-Making). In particular, if such a system is opaque for one or all of the reasons set out above, it is difficult to comply with the imperative to provide public reasons for most judicial decisions. If a system does not give any reasons, the value of open justice is undermined (the first obstacle). If a system gives 'true' reasons for the decision, (the pathway that the program took to go from the input data to the output decision), then those reasons would be incomprehensible to most if not all persons attempting to interpret them (the second obstacle). If a system gives 'comprehensible' reasons (a gloss on the 'true' reasons to provide interpretability to parties to the disputes), are those reasons are in fact meaningful or reflective of the decision-making process (the third obstacle)?

Ensuring that a system gives 'reasons' will facilitate understanding of the inputs and outputs of an AI system, but the internal operations of the decision-making process are less well understood. Even where a transparent system would allow judges and litigants alike to understand the 'process' which resulted in a decision, the decisions of the system may not be explainable.[165] Even some who create AI systems are unable to track their program's reasoning in the sense of understanding *why* the system produced a particular output.[166]

THINGS TO CONSIDER – Questions for courts

1. Are there ways that AI tools can *enhance* open justice by providing more people with practical access to court proceedings and decisions?
2. Are litigants and members of the public aware that AI is being used in this way? Are litigants and members of the public told *how* AI is being used (for example, to support or replace a human decision)?
3. Where AI is used to automate processes or support decision-making that would normally require reasons or explanations be given to those affected, is the functioning of the tool opaque to decision-makers and those impacted for reasons including as a result of:
   i. lack of disclosure by the provider of the AI system;
   ii. contractual promises to keep information confidential;
   iii. privacy of personal information in training data;
   iv. information being provided in a form that cannot be understood by the relevant audience (for example, as computer code);
   v. information being provided about the functioning of the system that is too complex for a functional understanding of the reasons for a given output.
4. Where automated 'reasons for decision' are produced, are they sufficiently comprehensible and an accurate representation of the operation of the AI system (bearing in mind the potential tension between these two objectives)?

Press 2020) 9 14–15; Hannah Bloch-Wehba, 'Access to Algorithms' (2020) 88 *Fordham Law Review* 1265, 1270; Deven R Desai and Joshua Kroll, 'Trust but Verify: A Guide to Algorithms and the Law' 31 *Harvard Journal of Law & Technology* 1, 10; Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) 31 *Harvard Journal of Law & Technology* 862–70; Matthew G Rowland, 'Technology's Influence on Federal Sentencing: Past, Present and Future' (2020) 26 *Washington and Lee Journal of Civil Rights and Social Justice* 565, 602 ('Its logic can be made more transparent and easier to adjust than human thinking').

163 Richard M Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22 *Stanford Technology Law Review* 242, 264–5.
164 Danielle Keats Citron, 'Technological Due Process' (2008) 85(6) *Washington University Law Review* 1249, 1252.
165 Re and Solow-Niederman (n 163) 262–267; Andrew D Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) 87 *Fordham Law Review* 1085.
166 Leah Wisser, 'Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing' (2019) 56(4) *American Criminal Law Review* 1815.

4.2 Judicial Accountability

Accountability can be undermined by AI tools because judges are likely unable to provide, or explain, the reasons an AI system whose outputs are used in making a decision has come to its conclusion.[167] AI tools might therefore decrease the accountability of judicial officers where the tools they rely on are opaque for the reasons outlined in the previous section. Traditional accountability mechanisms, including the right to appeal and the judicial obligation to give reasons, are less effective if judges only know the outputs of an opaque AI system.

The first significant obstacle to judicial accountability posed by AI tools comes from the decision by suppliers to keep secret information about those tools, requirements imposed on others through contractual and equitable confidentiality requirements (see discussion of COMPAS and other prediction tools in section 4.1 Open Justice). While the law manages general disputes relating to trade secrets in the face of litigants seeking information, courts should bear in mind the importance of judicial accountability when agreeing to purchase AI systems about which information is not publicly available or when agreeing to keep confidential information about the systems they use.

However, the same concern arises in relation to, for example, documents discovered and classified by way of TAR (see section 3.1 Technology Assisted Review and Discovery), which can form the foundation of a proceeding, and yet the judge or parties may have no real understanding of the method employed by a vendor's software. As Magistrate Judge Peck observed in *Da Silva Moore v Publicis Groupe*, the effectiveness of TAR could instead by assessed by reference to the precision and recall of that system. The court is also able to rely on the solicitor's duty to the court to ensure that their client makes proper discovery. The solicitor is therefore obligated to comprehend the operation of the particular form of TAR that is employed.

The secret nature of many AI systems means that judges, in addition to parties, will be unaware of the way in which outputs were generated. The issue will also affect appellate decisions where AI tools were used in reaching a first instance decision. To combat this concern, the Estonian small-claims court, which is intended to consist of an AI judge deciding lower-level contract disputes (see section 3.5 Automated Decision-Support and Decision-Making), would be subject to a right of appeal from the automated judicial decision.[168] The way in which this appeal would function is unclear. For example, would the human appeal judge reassess the same fact matrix in relation to the same regulation to be applied, or is the appeal limited to a question of whether the AI system itself is prone to error? Further, if the human appeal judge were to determine that the AI system had fallen into error, who would be accountable for that error? Would the system developer be required to remedy any failure correctly to decide a dispute, and if so, does that call into question issues of equality before the law as some litigants will be faced with an AI system different to that deployed in previous cases?

Where courts or other public institutions contract or commission AI tools for the purpose of delivering public services, and in particular where such public service necessitates a high degree of transparency as is required for judicial purposes, accountability enhancing features should be included in the terms of the contract or commission.[169] Alternatively, systems that incorporate explanation pathways that act as an intermediary between the source code of the program and the communication of the process with parties and judges can be developed so that the computer code can remain secret while still providing for accountable decision-making.[170] Where AI systems are used in the courtroom, a report should accompany its use which provides a sufficient explanation to the judge and parties, appropriate for the context of its use. Where such safeguards are not in place, courts should be wary of using AI systems' outputs in ways that affect the rights and obligations of individuals in circumstances where they have no real prospects of understanding or challenging the operation of the system.

Some jurisdictions have sought to address lack of accountability through specific legislative instruments. For example, the EU's General Data Protection Regulation (GDPR) prohibits decisions 'based solely on automated processes, including profiling, which produces legal effects concerning [a person] or similarly significantly affects him or her'.[171] Profiling is 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements'.[172]

167 Selbst and Barocas (n 165) 1088.
168 Tangermann (n 100).
169 Bloch-Wehba (n 162) 1308; *Substitute House Bill 1655 2019*.
170 Finale Doshi-Velez et al, *Accountability of AI Under the Law: The Role of Explanation*, Working Draft, 2017, 16–17.
171 *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ L 119/1.
172 'Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance)' *OJ L* 95 <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679> art 4(4).

A 2017 report by the UK House of Lords All-Party Parliamentary Group on AI explained that the use of AI in decision-making must 'coexist with accountability frameworks', so that '[a]ccountability and liability frameworks… [are] instilled to form structured guidelines for who/what is accountable for what. This will prevent leeway to interpretation and social mistrust'.[173] If the courts increasingly implement AI tools which judges may not themselves understand, how can they remain accountable to the public?

It is also worth considering the impact of RaC projects on judicial accountability. These projects are focussed on improving administrative decision-making and compliance tools, not judicial decision-making. Nevertheless, judges will need to consider questions such as the status of machine-consumable versions of legislation and the administrative law principles that apply to automated decisions. Judges will also play an important role in setting boundaries for automation in administrative contexts. For example, those involved in RaC projects recognise that discretionary decision-making power cannot be exercised by machines and that replacing discretionary powers with strict rules may result in arbitrary, irrational or unfair outcomes in certain cases.[174] However, there is potential that such known limitations may be ignored if governments seek to expand the use of automated decision-making in government, for example by replacing the exercise of a discretion with a prediction of how that discretion is likely to be exercised (see section 3.3 Prediction of Litigation Outcomes). Developments in machine learning may provide further temptations to do this. Judges thus have a strong role to play in ensuring the accountability, not only of judicial decisions, but of the use of artificial intelligence and automation in public decision-making more broadly.

THINGS TO CONSIDER – Questions for courts

Where an AI system may be deployed by a court or tribunal in a manner that might impact on the rights and interests of litigants or others:

1. What are the current mechanisms to ensure accountability in the relevant context (for example, reasons for decision and rights to appeal)?
2. Are these impacted by the proposed deployment, for example:
   a. Might reasons for decision simply refer to the output of an opaque AI system to explain the decision or a component thereof?
   b. Is information publicly available as to how outputs of the system are applied or moderated by humans?[175]
   c. Will any appeal process be able to override an erroneous system output?
   d. Who is required to answer for and correct a system in the context of a specific or systemic error?
3. How might accountability of fully or partly automated decision-making be enhanced?
   a. Would greater transparency of AI components of the decision-making process be useful or sufficient? If so, is this practically possible and contractually permitted?
   b. Can accountability and transparency of AI systems be improved through better procurement practices (including tailored requirements as to technical specifications and transparency)?
   c. Has independent testing been conducted to verify the system's overall performance and reliability? Are AI systems independently evaluated to ensure they meet important criteria (depending on the context of deployment) such as accuracy and non-discrimination (see section 4.3)? Is the evaluation published (ideally following peer review) and available to decision-makers and those impacted?
4. How secure is the system? Might its outputs be corrupted by malicious actors?

173 House of Lords, Committee Room 2, *Ethics and Legal in AI: Decision Making and Moral Issues A Theme Report Based on the 2nd Meeting of the All-Party Parliamentary Group on Artificial Intelligence [APPG AI]*. (No 2nd Meeting of APPG AI, Big Innovation Centre, 23 March 2017) 6 <http://www.appg-ai.org/wp-content/uploads/2017/03/themereport2_27mar.pdf>.
174 Bennett Moses, Boughey and Crawford (n 152).
175 Darren O'Donovan, 'Evaluating Automation: The Need for Greater Transparency' in Janina Boughey and Katie Miller (eds), *The Automated State: Implications, Challenges and Opportunities for Public Law* (Federation Press, 2021) 31, 48.

4.3 Impartiality and Equality Before the Law

Judicial values of impartiality and equality before the law intersect with AI tools in three discrete ways. First, AI systems may be used in biased ways in relation to people who are historically marginalized and/or structurally discriminated against (section 4.3.1 Bias and Discrimination in AI Tools). Second, AI systems may depart from fundamental principles of equality before the law by treating different cases as identical. Third, AI systems can identify instances of bias by members of the judiciary, particularly in areas such as race, gender or age (section 4.2.1).[176] By deploying AI systems to generate judicial metrics that identify areas of bias (section 4.3.2 Analytics to Measure Judicial Bias), AI tools could support the judicial value of equality before the law. We discuss these three different interactions in more detail.

4.3.1. Bias and Discrimination in AI Tools

Equality before the law sits uncomfortably with data-driven decision-making, such as that used in machine learning. Differential treatment based on *who someone is* as opposed to simply *what that person did* is almost universal. Variables in COMPAS include not only age and gender, but also information such as whether the defendant's parents still live together. Even where the use of such variables increases the accuracy of predictions (as it was argued for gender in COMPAS), that does not imply that treating people differently because of those variables is consistent with the principle of equality before the law.

Humans are also prone to bias. Sometimes this is overt but, more often, it is unconscious.[177] A collection of psychological studies suggest that humans, including decision-makers, rely on heuristics and cognitive short cuts, and are susceptible to effects such as decision fatigue.[178] The question is thus not whether humans or AI systems are more or less impartial, but rather what systems need to be in place to ensure that decisions are not biased in unacceptable ways. For judges and juries, there are rules about evidence as well as appeal pathways. For machines, we should ask about *what* is being optimised and *how* these choices are made.

Bias in human systems can be duplicated or enhanced in automated systems in different ways. We illustrate five of them in this guide. First, in situations where the training data is not representative or is generated through biased human action (for example, arrest data where police target certain groups). This appears to be the most difficult example of bias to overcome in current machine learning methods – most obviously, ProPublica's 2016 investigation indicated that African American defendants were more likely to receive a false positive COMPAS risk assessment score, whereas white defendants were more likely to receive a false negative COMPAS risk assessment score (see section 3.4 Criminal Sentencing and Risk Assessment Tools). Another example comes from attempts to predict property settlements in family law litigation – how will such systems contend with historic data reflecting gendered patterns of work? Even where such known biases are managed, system designers will need to contend with other ways in which bias can be introduced. As Bell has observed, where training data comes from court databases, it represents an atypical minority of family separations as most are not resolved through the courts.[179]

Second, systems may 'overfit' training data that is not representative of a broader population. Consider technologies such as intelligent speech processing, which could be used to replace court reporters and keep a live transcript of court proceedings.[180] This emergent bias arises from users' interaction with specific populations so that the system learns or adapts to particular groups and their responses over time.[181] If the models are trained predominantly on, for example, English-speaking, comparatively wealthy, non-minority datasets, it may then have greater difficulty in analysing and interpreting accents or dialects which do not comport with that community.

Third, humans may over-rely on outputs of AI systems, assuming that they are objective or 'scientific'. AI systems bring with them what is known as automation bias or 'algorithmic authority', which has been described as the 'decision [by human decision-makers] to regard as authoritative an unmanaged process of extracting value from diverse, untrustworthy sources, without any human'.[182] Replacing or supplementing discretion with AI systems is delegating 'some of our moral responsibility',[183] and yet the decisions of AI systems may be perceived as more 'reliable' and 'trustworthy'. This is exemplified by decisions where judges have overridden their own initial thoughts due to AI outputs, for example Judge Babler who increased a defendant's sentence of imprisonment based on COMPAS

176 Ozkan Eren and Naci Mocan, 'Emotional Judges and Unlucky Juveniles' (2018) 10(3) *American Economics Journal: Applied Economics* 171.
177 You can take a test to explore your implicit biases at: https://implicit.harvard.edu/implicit/takeatest.html.
178 See S Danziger, J Levav and L Avnaim-Pesso, 'Extraneous Factors in Judicial Decisions' (2011) 108(17) *Proceedings of the National Academy of Sciences* 6889; Legg and Bell (n 9) 96-97.
179 Felicity Bell, 'Family Law, Access to Justice, and Automation' (2019) 19 *Macquarie Law Journal* 103, 118.
180 Ward (n 131).
181 Rebecca Crootof, '"Cyborg Justice" and the Risk of Technological–Legal Lock-In' (2019) 119(1) *Columbia Law Review* 1, 4; See further Gary Edmond and Kristy A Martire, 'Just Cognition: Scientific Research on Bias and Some Implications for Legal Procedure and Decision-Making' (2019) 82(4) *Modern Law Review* 633.
182 Clay Shirky, 'A Speculative Post on the Idea of Algorithmic Authority' <www.shirky.com>.
183 Lucas D Introna, 'Maintaining the Reversibility of Foldings: Making the Ethics (Politics) of Information Technology Visible' (2007) 9(1) *Ethics and Information Technology* 11, 23 ('Maintaining the Reversibility of Foldings').

risk scores.[184] Ultimately, reliance on AI systems could *legitimate*, rather than *legitimise*, judicial decisions deeply embedded in discrimination and bias. Moreover, a judge, having been informed of a COMPAS risk score, may subconsciously delegate the difficult task of sentencing to an AI system which he or she trusts implicitly.[185] At the very least, we could expect risk scores to generate a framing or anchoring effect.[186]

Fourth, there are particular concerns that arise in the context of continuous machine learning (see section 2.6 Machine Learning). One concern with continuous learning is the possibility of feedback loops, where decisions taken through the operation of the system influence how it is trained over time. For example, consider a system that estimates the risk of a convicted and incarcerated individual re-offending after their release. Assume it is used in making parole decisions. If one continuously trains the system based on the behaviour of individuals after they are released, this data will be impacted not only by the tendencies of those individuals but the decisions that have been made about them based on outputs of the system being continuously trained. An error resulting in certain groups receiving a higher risk score than their 'true' risk score may lead to lengthier sentences for individuals in that group, which itself increases the likelihood of reoffending. The machine learning system will then 'learn' from that instance of reoffending and may designate an even higher risk score to future offenders in that group. There are ways of mitigating against this, for example by using skewed sampling or synthetic data in continuous learning to reverse the impact of such feedback loops, but they are not always implemented.

Fifth, a lack of interpretability creates a risk that more financially capable parties will be able to gain a greater level of understanding of technological systems, by hiring experts in the field, and consequently 'game' the judicial process. This asymmetry would also undermine equality before the law.[187]

Principles of impartiality and equality before the law require not only that like cases are treated alike, but also that different cases are treated differently.[188] AI systems, whether based on a pre-programmed logic or machine learning, draw on specific inputs. Rarely are systems designed with an 'other' category that would allow for consideration of unanticipated factors. An expert system will only consider factors that were contemplated at the time it was programmed. Similarly, machine learning may cluster or classify together cases that ought to be treated differently simply because the type of fact that makes the cases different was not built into its model. In such situations, equality before the law can be denied because relevant distinctions are not drawn.

4.3.2. Analytics to Measure Judicial Bias

AI systems can also be deployed in the courtroom for analysis of judicial decisions. For example, in cases of high commercial value, past patterns of conduct may shape the way in which a case is presented as lawyers 'craft' arguments tailored to appeal to certain judges, producing an echo chamber in which each application of the data generates confirmatory data.[189] As described in section Prediction of Litigation Outcomes, evidence of supposed judicial bias on the basis of AI analysis of the outcome of proceedings has been an active area of research. However, as with any application of data analyses to individuals, there is the risk that individual differences or nuances of a case are overlooked in pursuit of machine-generated, and machine-recognisable, patterns.

Statistical evidence which goes to the bias or partiality of particular judicial officers has not yet met the threshold for proving apparent or actual bias for the purpose of a recusal application. In Australia, statistical evidence has been rejected as lacking probative value, and held not to reach the high threshold of actual or apparent bias.[190] However, such statistical analyses can indicate tendencies among judges to rule in particular ways, and presentation of such information may fuel public criticism of judges.

Further, as recent events in Australia have made clear, researchers and members of the public must be cautious of the way that such data is interpreted. Deference to statistical figures, particularly when taken out of context, can be damaging when attacks on judicial impartiality corrode judicial independence. Partly in fear of circumstances in which the impartiality of a judicial officer, and thereby the judiciary in general, is undermined, in 2019 France passed a law which specifically proscribed the conduct of correlating the 'identity data of magistrates and members of the registry... with the object or effect of evaluating, analysing, comparing or predicting their actual or supposed professional practices'.[191]

184 Carlson (n 79).
185 'Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing', *UCLA Law Review* (19 February 2019) <https://www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/> ('*Injustice Ex Machina*'); who cites examples of evacuees blindly following malfunctioning robots in emergency situations, and seasoned radiologist relying on faulty diagnostic aids despite having expert knowledge to override those diagnostics.
186 Todd McElroy and David Dickinson, 'Thoughtful Days and Valenced Nights: How Much Will You Think about the Problem?' (2010) 5 *Judgment and Decision Making* 516.
187 Richard M Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22 *Stanford Technology Law Review* 242, 264, 266–7.
188 *Wong v The Queen* [2001] HCA 64; 207 CLR 584 [65]; *Green v The Queen* (2011) 244 CLR 462, 472–3 [28].
189 Pasquale and Cashwell (n 64).
190 *BDS17 v Minister for Immigration and Border Protection* [2018] FCA 1683.
191 *Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice (1)* (France) JO, 23 March 2019, art 33; Artificial Lawyer, 'France Bans Judge Analytics, 5 Years In Prison For Rule Breakers', *Artificial Lawyer* (Online, 4 June 2019) <https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/>.

The use by a judiciary of AI systems has a complex relationship with the promotion of the judicial values of impartiality and equality before the law. Tools such as COMPAS and other risk assessment programs can entrench, rather than neutralise, instances of bias in the judiciary. As Chief Justice Bathurst AC stated in his 2021 Sir Maurice Byers lecture,[192] '[a] machine appears impartial: it weighs up the data before it with ruthless dispassion and is unaffected by emotion'. However, as he also notes, the compatibility with the judicial function of impartiality is only 'superficial'.[193]

Australia is currently considering how AI and other technological tools can impact on judicial impartiality in ongoing federal reform;[194] and we hope that the potential for AI tools to undermine judicial impartiality will lead to some tangible policy action. If AI systems are used to determine the rights and interests of an individual, particularly in circumstances where that individual risks their liberty, we must not simply replicate human bias, but should design systems in ways that enhance our commitment to equality before the law. This should start with a requirement for open and accessible technologies to be used in courts, as opposed to those subject to confidentiality requirements. Further, AI tools can be used to detect instances of human bias which portend inequality before the law. This may allow the judiciary to move beyond the 'impartial enough' status quo'.[195]

THINGS TO CONSIDER –
Questions for those considering the use of AI systems in courts, tribunals and registries

1. Has the AI system been evaluated to determine whether its outputs might be biased in problematic ways? There are different ways of measuring 'fairness', so a system might be fair on one metric but unfair on another – what is fair will depend on context in addition to relevant legal requirements.
2. In the context of machine learning, what training data was used? Might this be skewed because of:
   a. bias in historic decisions that impact on the data? For example, when police target particular populations, this can lead to skewed data in crime databases.
   b. bias in historic 'facts'? For instance, average male and female incomes have differed for reasons unrelated to role or performance.
   c. overrepresentation or underrepresentation of particular populations? This can occur due to historic marginalisation.
   d. feedback loops, where the data collected is impacted by decisions influenced by the outputs of an AI system. For example, chances of re-offending are impacted by decisions made in relation to parole which are based on automated risk assessment tools.
3. In the context of machine learning, does the training data include variables that might be proxies for categories protected by discrimination legislation?
4. In the context of machine learning, does the system change over time (continuous learning) and will this result in unfair differences between decisions made at different times?
5. Does the AI system make assumptions that may be no longer be valid? Is there a possibility of overriding the AI system in the event of unanticipated factors?
6. Are those using the system aware of its limitations and trained to avoid overreliance?
7. Is information about an AI system made available to all, giving those affected an equal ability to understand its outputs and appeal as required?

In addition, courts and tribunals should consider the use of AI systems for measuring decisions made by human judges and officers with caution.

192 The Hon TF Bathurst AC, 'Modern and Future Judging' (at the Sir Maurice Byers Lecture, 3 November 2021) [54].
193 Ibid.
194 ALRC, 'Review of Judicial Impartiality' (Webpage, 11 September 2020) <https://www.alrc.gov.au/inquiry/review-of-judicial-impartiality/>.
195 Avital Mentovich, JJ Prescott and Orna Rabinovich-Einy, 'Are Litigation Outcome Disparities Inevitable? Courts, Technology, and the Future of Impartiality' (2020) 71(4) *Alabama Law Review* 893, 897; See also Charles Gardner Geyh, 'The Dimensions of Judicial Impartiality' (2014) 493 *Florida Law Review* 498, 492, 510.

4.4 Procedural Fairness

Procedural fairness, also called natural justice is 'central to the rule of law and includes receiving notice of a claim and the opportunity to be heard'.[196] Though 'fairness' more generally has been a concern discussed in relation to AI applications, procedural fairness is a distinct issue with a particular meaning where there is an exercise of judicial power.

The opportunity to be heard before a judicial decision is made was described by Justice Heydon as occurring by way of a hearing:[197]

> A hearing takes place before a judge at a time and place of which the moving party has given notice to the defending party. At it both parties have an opportunity to tender evidence relating to, and advance arguments in favour of, the particular orders they ask for. This aspect of the rules of natural justice pervades Australian procedural law.

Procedural fairness may also include the parties being given an opportunity to call their own witnesses and to cross-examine the opposing witnesses.[198] An opponent may not advance contentions or adduce evidence of which a party is kept in ignorance.[199] The impact of the use of AI tools on procedural fairness will depend on the nature of the tool and the context of its use. At one extreme, the idea of an AI system *making judicial decisions* based on fixed inputs would deny natural justice. On the other hand, many AI tools operate outside of the exercise of judicial power and therefore do not impact procedural fairness in the manner defined above. Where AI is used to generate evidence that cannot be easily challenged (as may occur for risk assessment tools) or to prevent cases coming on for a hearing through automated triage, then procedural fairness is clearly implicated. In all of these examples, it is worth considering whether litigants are truly heard. This will depend on how AI systems are designed and deployed – do litigants shape the inputs to an extent that one can say that the system has truly factored in evidence presented and listened to parties' arguments?

Procedural fairness concerns also arise through the use of risk assessment tools. One of the claims made by Eric Loomis in his appeal (see section 3.4 Criminal Sentencing and Risk Assessment Tools) was that his right to due process was infringed due to the court's reliance on a risk assessment which he was unable properly to challenge. The inability to question the risk assessment was because it was not possible to know how the COMPAS tool had 'weighted' the different inputs. In that case, the Court rejected the due process argument. Though agreeing that the use of such a tool raised due process concerns, it held that cautious and selective use was acceptable.

Finally, automated decision-making would also likely breach rights to procedural fairness, dependent on how it was to be used, whether parties consented to its use, or not, and what rights of appeal followed (see section 3.5 Automated Decision-Support and Decision-Making). For example, extrapolating on the operation of the EXPERTIUS system in Mexico described above, if it involved an exercise of judicial power then allowing for the filing of evidence but not permitting argument or the challenging of the material relied on to determine the pension would be unacceptable in jurisdictions such as Australia. Consequently, at present, most AI systems do not seek to adjudicate the outcome of disputes.

Procedural justice is also important in the context of appeals from decisions influenced by AI systems (also discussed under section 4.3 Impartiality and Equality Before the Law). As highlighted through the discussion of the facts in *Hemmett v Market Direct Group Pty Ltd* [No 2],[200] those in a position to overturn an automated decision need to understand the operation of the system concerned to assess its compliance with relevant requirements. According to the NSW Ombudsman, this requires maintenance of a register of all systems in use, with dated descriptions of version changes and cross-references to any changes in law or policy that necessitated those changes.[201] Historic versions of systems should also be archived.

196 Michael Legg, 'The COVID-19 Pandemic, the Courts and Online Hearings: Maintaining Open Justice, Procedural Fairness and Impartiality' (2021) 49 (2) *Federal Law Review* 161, 169 ; *K-Generation Pty Ltd v Liquor Licencing Court* (2009) 237 CLR 501, [48] (French CJ); *International Finance Trust Co Ltd v New South Wales Crime Commission* (2009) 240 CLR 319, 354 [54] (French CJ) and *HT v The Queen* [2019] HCA 40, [17] (Kiefel CJ, Bell and Keane JJ), [64] (Gordon J).
197 *International Finance Trust Co Ltd v New South Wales Crime Commission* (2009) 240 CLR 319, 379 [141].
198 *Al Rawi v Security Service* [2011] UKSC 34; [2012] 1 AC 531, [13].
199 Ibid [12].
200 [2018] WASC 310.
201 Ombudsman New South Wales, *The New Machinery of Government: Using Machine Technology in Administrative Decision-Making* (29 November 2021) 48.

THINGS TO CONSIDER – Questions for courts

1. What are the real and perceived impacts on procedural fairness if a particular AI system is deployed?
   a. Are litigants given a real opportunity to have their case and evidence heard and considered, or do limitations on system inputs and operations affect this? For example, can system outputs be challenged where litigants feel that the inputs were in error or that the system fails to take account of relevant factors? Is there room for argument in how system outputs impact on final decisions?
   b. Will litigants feel that they have been heard, so as to feel satisfied (if not happy) with the outcome and retain trust in the justice system?
2. Are courts retaining and making available sufficient information on systems in use to ensure that rights to appeal against decisions made or influenced by such systems are preserved. In particular:
   a. Is there a register of all systems in use?
   b. Are the dates of and reasons for version changes retained?
   c. Are historic versions of systems archived?

4.5 Access to Justice

One of the great promises of AI in law is that it will enhance access to justice. Access to justice refers to an ability learn about legal issues and seek redress for legal problems. This may involve dispute resolution processes outside of court but also the right to approach a court. Susskind and other scholars argue that the legal profession should seek to automate more tasks, including through the use of AI, in order to drive down the costs of legal services thus making them more accessible.[202] While those questions are beyond the scope of this guide, access to justice considerations are also relevant to how courts engage with the opportunities presented by AI.

ODR processes may facilitate access to justice by incorporating targeted information about the law relevant to a person's dispute and then streamlining processes and allowing dispute resolution steps to be done remotely (see section 3.2 Automated Online Dispute Resolution). For example, eBay's Resolution Centre was set up to manage a high volume of low value disputes, where buyers and sellers might be located in different jurisdictions, in a largely automated way.[203] Ease of access and efficiency means that parties to a dispute are more likely to make use of this ODR process. This is the case, too, with online courts for small claims matters such as that proposed in England and Wales and operating in British Columbia. Generally, only part of the court process is automated, such as the intake process where a person receives assistance to file their claim.

Automation can also be used to increase access to justice through developing chatbots which can answer questions and direct users to better-tailored information. Some US courts are using chatbots to address commonly asked questions and therefore reduce calls to court personnel. For example, the Los Angeles Superior Court developed a chatbot for this purpose in June 2020:

> To get the bot up and running quickly and efficiently, the court designed it along the same lines as the chatbot used to order Domino's Pizza. The chatbot uses preliminary or guiding questions to lead users to the right answers from a knowledge base of 100 questions based on user guides and FAQs.[204]

These types of question-and-answer system might then segue into the commencement of a claim – for example, the chatbot developed by Joshua Browder (Do Not Pay), which generated a letter for the user to challenge their parking infringement.[205]

However, 'access to justice' is not only about 'access' but also about 'justice'. Triage tools (3.7 Triaging and Allocation of Matters) could be seen as restricting access to justice if they either impose a longer time frame on or prohibit some people from applying to a court altogether. For example, it was suggested that the machine learning system built by Aletras et al to predict decisions of the European Court of Human Rights (see section 3.3 Prediction of Liti-

202 For example, Richard Susskind and Daniel Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford University Press, 2015); Richard Susskind, *Transforming the Law: Essays on Technology, Justice and the Legal Marketplace* (Oxford University Press, 2000).
203 Legg and Bell (n 9) 140-143.
204 La Roque-Doherty (n 113).
205 'Save Time and Money with DoNotPay!' <https://join.donotpay.com>.

gation Outcomes) could be used to triage matters and prioritize those most likely to succeed.[206] This could impinge on a person's right to be heard (see previous section on procedural fairness). Also, making decisions with significant impact on people's lives without engaging with them through a human process may lead to dehumanisation and failure to treat people with dignity.[207] Similarly, human experience and discretion is central to most judicial decisions and cannot be meaningfully exercised by any known AI systems.[208] More broadly, losing judges' emotion, morality, indeterminacy and creativity would fundamentally change what justice looks like.[209] It is thus crucial to ensure that the cost of enhanced access or greater efficiency (discussed below) is not the value in the system itself.

THINGS TO CONSIDER– Questions for courts

1. Are there ways in which AI can improve the operation of courts to enhance access to justice, including through reducing delays?

2. What are the broader implications of employing such tools?

3. Might the deployment of an AI system, particularly in the context of triage tools, reduce access to justice for some?

4.6 Efficiency

Efficiency – the saving of cost and time – is perhaps the most compelling reason for the use of AI tools in justice systems and courts. It is recognised that there is an existing tension between the need to ensure procedural fairness and justice, yet also proceeding in an efficient manner.[210] The link between efficiency and AI is clear; as judges themselves are both an expensive and limited resource,[211] automated decision-making has been suggested as a cheap, fast and scalable alternative.[212] However, efficiency comes with 'implied strings'[213] or 'trade-offs',[214] including risks for other judicial values. Most technology projects within the courtroom are aimed at increasing efficiency and minimising expense – usually by saving judges, registrars, officers and litigants cost and time. Many of the basic tools discussed in section 3 are among the most important for judicial work and, with some exceptions discussed below, largely uncontroversial. Improving the ease of document retrieval[215] and the use of e-filing systems (see section 3.6 Automated E-Filing)[216] to streamline and synchronise the operation of registries and judges are welcome developments. Efficiency gains will be greatest for administrative steps incidental to the exercise of the judicial function, such as automated e-filing, triaging and allocation of matters, and automated transcription services. These applications are only efficient if they operate as intended and exhibit a high degree of accuracy. For example, limitations of AI transcription services include the need to format output documents, accuracy, and difficulties associated with having multiple voices in the courtroom. Additionally, automated transcription services are often less able to contextualise statements than a human listener.[217]

Systems such as TAR (see section 3.1 Technology Assisted Review) should make litigation involving voluminous discovery more efficient for the parties and their lawyers, but not necessarily so for the court. Indeed, TAR, and the use of risk assessment tools, could generate more work for courts as they increase the number of subsidiary issues that parties dispute between each other without lessening the court's existing workload. It would seem likely that in complex disputes, particular in the early stages of implementation of technologies like TAR which seek to increase the efficiency of disputes, there is likely to be a need for judge-led case management.

Likewise, automated ODR or automated decision-making are efficient insofar as they prevent matters coming to court that would otherwise have been litigated. If they simply become a first step, the workload of the court will not be altered, and in fact disputes may be made more complicated. If the decision of an automated system is appealed, the appellate decision-maker must review the information presented to them and make their own assessment of its veracity and weight in their consideration. Consequently, it is unclear whether the addition of this factor, perhaps difficult for the judge to understand, really makes for more efficient court proceedings.

206 Aletras et al (n 62).
207 Sourdin (n 11) 49, quoting Margot E Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92 *Southern California Law Review* 1529, 1542.
208 Meena Hanna, 'Robo-Judge: Common Law Theory and the Artificially Intelligent Judiciary' (2019) 29 *Journal of Judicial Administration* 22, 39–42.
209 Bathurst AC (n 192).
210 Michael Legg, 'Reconciling the Goals of Minimising Cost and Delay with the Principle of a Fair Trial in the Australian Civil Justice System' (2014) 33(2) *Civil Justice Quarterly* 157.
211 Crootof (n 167) notes that 'human judges are an inherently expensive and limited resource: They must prepare for years, they take time to decide cases, and they retire'.
212 Re and Solow-Niederman (n 163).
213 Jennifer Walker Elrod, 'Trial by Siri: AI Come to the Courtroom' (2019) 57 *Houston Law Review* 1083, 1093.
214 Sabine Gless, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2019) 51(2) *Georgetown Journal of International Law* 195, 231.
215 Sheryl Jackson, 'Court-Provided Trial Technology: Efficiency and Fairness for Criminal Trials' (2010) 39(3) *Common Law World Review* 219, 236-7.
216 'HMCTS E-Filing Service for Citizens and Professionals', *GOV.UK* (26 April 2021) <https://www.gov.uk/guidance/hmcts-e-filing-service-for-citizens-and-professionals>; 'Electronic Filing (CM/ECF)' (n 40); Federal Court of Australia (n 111).
217 Kim Neeson, 'Is AI Coming to a Legal Transcript Near You?', *LexisNexis Canada* (June 2019) <https://www.lexisnexis.ca/en-ca/sl/2019-06/is-AI-coming-to-a+legal-transcript-near-you.page>.

Care must be taken against attempts simply to improve efficiency in a strictly business sense, for instance by considering the judicial role as a typical ‘service’, and minimising the ‘cost’ associated with putting out the same ‘volume’ of decisions. There may be strong *cost-saving* reasons to make all hearings online or, perhaps, to limit all hearing lengths to five minutes. However, such attempts to maximise economy would undermine other judicial values such as that of judicial transparency and access to justice.

Generally, AI systems must function with a high degree of accuracy to reap any efficiency gains. However, this level of accuracy is not always achieved. For example, unless there is a means of checking the validity of inputs, automated e-filing may ultimately create more work if errors that would have been identified by a registry clerk are missed. Given the cost involved in creating AI tools to begin with, there ought to be sharply superior results for the attendant difficulties (eg opacity, possible bias) to be worthwhile.

Finally, Reichman et al, in an analysis of Israel’s Legal-Net system,[218] found that online systems implemented to promote economy and efficiency ‘nudged judges… to think of their role as part of the assembly line, the business of which is to produce dispute settlements under the law’.[219] This is concerning as it suggests an unintended consequence of AI use for judicial management is to alter the judicial role itself without sufficient thought.

THINGS TO CONSIDER – Questions for courts

1. Will a particular AI system enhance efficiency *as a whole*, bearing in mind that errors or failing to identify problems that would have been picked up by a human early may result in *inefficiency*? How will changes in overall efficiency be measured?

2. What harms will or might result from measures to improve efficiency, bearing in mind particularly the core judicial values discussed in this Part?

4.7 Interaction between AI and Judicial Values

This section explained how AI tools can impact on and interact with core judicial values. AI is an evolving field. Currently available tools are not sufficiently accurate, nuanced, and unbiased to replace judges or indeed many of the functions performed in court registries. In many contexts, there are no tools that would satisfy the standards required by the judicial values. However, AI systems can be, and are being, used appropriately by judges, court registries and parties. What is required is critical awareness of the circumstances in which AI is deployed. The questions posed throughout section 4 provide a useful place to start.

While rules as code is not a tool used directly in courts, it aims to enhance both transparency and efficiency in the administration of law by government. This may impact on how judges treat administrative decisions made by computer systems deploying code developed within rules as code projects.

The judicial values discussed in this chapter are not the only relevant consideration when determining whether a jurisdiction should implement a particular AI tool. The resolution of procedural issues or disputes over small claims may be less sensitive than criminal proceedings. Of course, judicial values remain important even in less critical contexts and even small claims can have a significant impact on the lives of vulnerable people.[220] Nevertheless, the closer the proposed use is to the core of judicial decision-making, the more caution is required.[221]

218 Reichman, Sagy and Balaban (n 119).
219 Ibid 635.
220 Sourdin (n 11) 235.
221 Joe McIntyre and Anna Olijnyk, ‘Public Law Limits on Automated Courts’ in Janina Boughey and Katie Miller (eds), *The Automated State: Implications, Challenges and Opportunities for Public Law* (The Federation Press, 2021) 89.

5. A Way Forward

AI is becoming popular in courts and tribunals across the globe. AI systems range from simple practices such as automated e-filing of documents to the complexity surrounding determining the likely risk that an offender will reoffend. The tension between the need for a judiciary to remain flexible as technology evolves and the normative principles of consistency and predictability in the resolution of justice is not a new phenomenon. In deciding whether any particular tool should be used in courts, members of the judiciary, as the persons overseeing the proper resolution of disputes, should be aware of the potential benefits which flow from the use of such technologies and their complex relationship with core judicial values.

The use of AI systems in the courtroom has consequences for open justice, accountability, impartiality and equality before the law, procedural fairness, access to justice and efficiency. Many of the issues raised in this guide, such as use of proprietary software trained on data which itself exhibits bias, can be mitigated through better specification of requirements in procurement and design processes. However, each AI system is different – it is necessary to ask whether, in relation to a *particular* system, there are *particular* concerns which could jeopardise the open, accountable, impartial, fair and efficient delivery of justice. Understanding the common AI terms and tools, together with the key areas of AI use in courts globally, can assist in examining the impact of AI on core judicial values on a case-by-case basis.

An important question beyond judicial values is whether the use of a particular AI system will be acceptable to litigants and members of the public. Her Excellency the Honourable Margaret Beazley AC QC asks whether individuals will feel they are treated fairly in their interaction with the legal system.[222] Will computer outputs warrant or receive the same respect as human judges?[223] Ensuring core judicial values are respected in the deployment of AI systems is necessary but not sufficient for public acceptance. Adherence to those core values must be communicated or seen to exist by the public.[224]

THINGS TO CONSIDER – Overarching questions about AI in courts and tribunals[225]

- *Why* is AI being used? What problem does it solve?
- Is the use of AI authorised in the context in which it is deployed?[226]
- *In what contexts* is AI being used, and is its use in those contexts appropriate? Does the context involve high stakes, vulnerable people, novel situations, or high levels of emotion?[227]
- *How* is AI being used? How can system requirements (through a procurement process) better fulfil its purposes and meet the needs of courts and tribunals, including in relation to core judicial values? How will the system be checked, tested and evaluated to ensure it meets those requirements?
- *Who is consulted* about the deployment of AI systems? Are all stakeholders including users and litigants included in decision-making about whether and how AI will be used?
- Will the use of AI impact on *public confidence in the judiciary*?[228]
- Will the use of AI in the courtrooms be accepted by the public?

222 Margaret Beazley, 'Law in the Age of the Algorithm' (Speech at the State of the Profession Address to New South Wales Young Lawyers, Sydney, 21 September 2017); cited in Sourdin (n 11).
223 See also Sourdin (n 11).
224 Legg (n 196) 181-183.
225 These higher level questions are based in part on those posed in Jeff Ward, 'When and How Should We Invite Artificial Intelligence Tools to Assist with the Administration of Law? – A Note from America' (2019) 93 *Australian Law Journal* 176.
226 Justice Melissa Perry, 'IDecide: Administrative Decision-Making in the Digital World' (2017) 91 *Australian Law Journal* 29; O'Donovan (n 175) 43.
227 Sourdin (n 11) 255–256.
228 On the importance of public confidence in the judiciary, see Hon Chief Justice Murray Gleeson AC, 'Public Confidence in the Judiciary' (2002) 76 *Australian Law Journal* 558.

APPENDIX 1: Survey of AIJA Members

INTRODUCTION

This survey is part of a collaborative project, 'AI Decision Making and the Courts', between the AIJA and a UNSW Law & Justice research team composed of Dr. Monika Zalnieriute, Prof. Lyria Bennett Moses, Jacob Silove, Prof. Michael Legg and Dr. Felicity Bell.

The project seeks to facilitate the preparation of a report for AIJA members, and judges in the Asia-Pacific region more generally, setting out:

- The key challenges and opportunities that automated decision-making presents for courts and judges;
- The different techniques falling under the umbrella of 'Artificial Intelligence' (AI) and their affordances and limitations;
- Examples of different contexts in which these techniques have been used in courts, both in Australasia and globally, together with a discussion of important issues arising in those contexts;
- Examples of judicial responses to such techniques, drawing on legislation, case law and rules in jurisdictions including the USA, UK and the EU.

The survey will enable the research team to identify the areas of interest of members of the AIJA, to ensure that the report is tailored to suit those interests. This will include gaining an appreciation of the particular judicial values that legal decision-makers are most concerned with in relation to AI technology in the courtroom, as well as the safeguards that the report should explore.

Following the survey, the intention of the research group is to conduct an extensive review of the use of AI tools by the judiciary in Australasia. Additionally, a roundtable will be organised to discuss report findings and to seek further feedback on the draft guide from AIJA members.

I. IDENTIFYING QUESTIONNAIRE PARTICIPANT

This section will be used to better appreciate the levels of understanding, interest and concern in the space of AI and the judiciary across different judicial and non-judicial institutions.

1. To which court/s or tribunal/s are you currently appointed?
   a. Free text
2. In what year were you first appointed to an Australasian court or tribunal?
   a. Drop down, year

II. CURRENT KNOWLEDGE OF TERMINOLOGY AND ISSUES IN THE AI SPACE

This section of the questionnaire will aim to establish your current level of awareness of the way in which automated systems, such as AI, work. The responses received from this section will determine the level of detail and complexity that the final report of the joint AIJA - UNSW report will incorporate when providing an understanding of the technical aspects of AI.

1. How confident would you feel in defining the following terms:
   a. Automation;
      i. Scale, 1 (not confident) – 5 (very confident)
   b. Machine learning;
      i. Scale, 1 (not confident) – 5 (very confident)

c. Machine learning lifecycle;
    i. Scale, 1 (not confident) – 5 (very confident)

d. Artificial intelligence or AI;
    i. Scale, 1 (not confident) – 5 (very confident)

e. Algorithms;
    i. Scale, 1 (not confident) – 5 (very confident)

f. Continuous learning;
    i. Scale, 1 (not confident) – 5 (very confident)

g. Technological 'black box';
    i. Scale, 1 (not confident) – 5 (very confident)

h. Deep learning and neural networks;
    i. Scale, 1 (not confident) – 5 (very confident)

i. Explainable AI (XAI);
    i. Scale, 1 (not confident) – 5 (very confident)

j. Supervised, semi-supervised and unsupervised machine learning;
    i. Scale, 1 (not confident) – 5 (very confident)

2. How important do you think it is for judges and tribunal members, for the purpose of their day-to-day work or for the future direction of courts and tribunals generally, to be more knowledgeable than at present about the terminology surrounding, and the operation of, AI?

    a. Scale, 1 (not important) – 5 (very important)

3. Please expand on the above answer (optional).

    a. Free text

4. When thinking about AI and other similar technologies, are there any phrases, concepts or implementations that come to mind which you believe the legal decision-makers in courts and tribunals ought to be more knowledgeable about?

    a. Free text

III. EVOLVING SKILLS OF THE JUDICIARY

This section of the questionnaire will aim to establish whether, in light of the answers given above, legal decision-makers should gain more knowledge or have greater skills in the technological sphere.

1. Would you have benefitted, in your role at your respective court/s or tribunal/s, from further education relating to emerging technologies such as AI?

    a. Scale 1 (not benefitted) – 5 (benefited greatly)

2. How important do you think the following types of education are for legal decision-makers in relation to emerging technologies and their interaction with the judiciary?

    a. Seminars targeting members of the judiciary
        i. Scale 1 (not important) – 5 (very important)

    b. Induction programs for new members of the court or tribunal
        i. Scale 1 (not important) – 5 (very important)

    c. Higher education (eg. courses at university)
        i. Scale 1 (not important) – 5 (very important)

3. Do you think it is the place of legal decision-makers, such as judges and tribunal members, to learn about and determine the appropriate level of implementation of AI in the courtroom? Why or why not?

    a. Free text

IV. CURRENT KNOWLEDGE OF IMPLEMENTATION OF AI IN THE JUDICIARY

This section of the questionnaire aims to determine the current level of knowledge of the ways in which AI has been implemented for judicial purposes.

1. Please briefly describe any AI tools that you are aware of which have been applied in a judicial context *in Australasia*.

    a. Free text

2. Please briefly describe any AI tools that you are aware of which have been applied in a judicial context *globally*.

    a. Free text

V. AREAS OF INTEREST – TECHNOLOGICAL INNOVATIONS

This section of the questionnaire aims to understand which areas participants are particularly interested in as a judge or member of their respective court/s or tribunal/s.

1. Which of the following examples of the use of AI systems in legal practice and legal decision-making would you be interested in learning more about (select all that apply)?

    a. Machine learning tools in discovery;

    b. Data-driven advice to clients regarding their likelihood of success or likely remedy in a matter;

    c. Prediction of judicial outcomes more broadly;

    d. Generation of litigation strategies;

    e. Natural language processing for translation and transcription purposes;

    f. Automated electronic filing;

    g. Triaging and allocation of matters;

    h. Judicial administration and judicial metrics;

    i. Simple cases involving few and defined elements being dealt with through automated systems;

    j. Automated decision-making or decision support in the context of small claims matters;

    k. Automated adjudication of matters with consent of the parties;

    l. Automated online dispute resolution;

    m. Risk assessment tools in sentencing;

    n. Estimation of damages or penalties based on automated formulae or data-driven tools;

2. Are there any other areas of legal decision-making and administration that you would be interested in learning more about in terms of possible implementations of AI systems and other sophisticated technologies?

    a. Free text

VI. PERCEIVED RISKS AND BENEFITS - AI AND THE JUDICIARY

This section of the questionnaire aims to understand the perceived values which could be benefitted or put at risk when AI is implemented in the judicial context.

1. Which of the following values and their relationship to AI in courts would you like to see analysed in the final project report?
   i. Open justice and transparency
   ii. Judicial accountability
   iii. Judicial independence
   iv. Judicial impartiality
   v. Procedural fairness
   vi. Access to justice
   vii. Economy and efficiency
   viii. Constitutionality
   ix. Other (please specify)

2. Are there any specific components of the above values that you believe ought to be considered in the report?
   a. Free text

VII. APPROPRIATE SAFEGUARDS

This section of the questionnaire aims to understand the key safeguards that you believe the final report should engage with in relation to a variety of possible implementations of AI in the legal decision-making context.

1. Which of the following safeguards do you think the report should examine as possible solutions to the potential risks posed by the use of AI in a judicial context?
   i. Ensuring that humans are involved in some stage of the automated processes.
   ii. Ability to appeal to a human decision-maker.
   iii. Preventing the underlying algorithm being covered by trade or state secrecy laws, or other intellectual property protections:
   iv. Ensuring that some form of reasons are given.
   v. Ensuring that the system developers are accountable for the output of an AI system.
   vi. Ensuring that the AI system produces demonstrably equal outcomes in relation to protected characteristics including age, disability, race, religion and sex.
   vii. Ensuring that there is no risk of tampering with or 'hacking' the AI system.

3. What, if any, other safeguards do you think the final report should examine in relation to the use of AI in a court or tribunal context?
   a. Free text

4. Do you have any other thoughts, comments or considerations about potential safeguards which can help direct the content of the report?

a. Free text

VIII. INFORMATION INCLUDED IN THE FINAL REPORT

a. Please let us know if you consider any other aspect of the intersection between technology and the law to be important to include in the final report, including or building on the questions in this survey.
   i. Free text

AIJA

The Australasian Institute of
Judicial Administration Inc.

www.aija.org.au