AI Decision-Making and the Courts

A guide for Judges, Tribunal Members and Court Administrators

Authors of original edition:
Felicity Bell / Lyria Bennett Moses / Michael Legg / Jake Silove / Monika Zalnieriute

Authors of current edition:
Lyria Bennett Moses / Michael Legg / Jake Silove / Monika Zalnieriute; with research assistance from Shahzeb Mahmood.

Joint research project between the Australasian Institute for Judicial Administration and UNSW Faculty of Law and Justice, through the UNSW Allens Hub for Technology Law and Innovation and the UNSW Centre for the Future of the Legal Profession.

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

T (61 2) 8099 2611
www.aija.org.au

The AIJA is a research and educational institute, which has been in existence since 1976. It is funded by the Commonwealth, states and territories and the New Zealand Ministry of Justice, and by subscription income from its membership. The principal objectives of the Institute include research into judicial administration and the development and conduct of educational programs for judicial officers, court administrators and members of the legal profession in relation to court administration and judicial systems. The AIJA is an Incorporated Association under the Associations Incorporation Act 1991 (ACT) and is governed by its Rules. The AIJA members include judges, magistrates, tribunal members, court administrators, legal practitioners, academic lawyers, court librarians and others with an interest in judicial administration.

The Institute also delivers programs of education to support those working in the field of judicial administration, including specific programs for court administrators, court librarians, magistrates and judges. The AIJA has also been involved in developing courses in specialised areas, including gender awareness, cultural awareness, court technology and case management.

Published June 2022

ISBN: 978-1-875527-60-1

Revised and republished December 2023

ISBN: 978-1-875527-66-3

Foreword

As my predecessor, The Hon. Justice Jenny Blokland, noted in her Foreword to the original publication of this AIJA guide (2022):

> *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.*

Recognising this, and given the rapid rate of developments in this field the AIJA commissioned an update to the guide in 2023.

The original guide provided a useful overview of various AI and automated decision-making tools relevant to courts and tribunals, and their potential impacts on core judicial values of open justice, accountability and equality before the law, procedural fairness, access to justice, and efficiency. It also outlined challenges and opportunities AI tools present and identifies primary interest areas and questions for courts and tribunals to consider when adopting certain AI tools, informed by a survey analysis and virtual roundtable discussions the research team conducted with AIJA members in 2021.

The 2023 guide includes updates to multiple sections, reflecting rapidly evolving case law, legislation and policies in this area; and the timely addition of new guidance on emerging technologies such as generative AI (e.g. Google Translate, ChatGPT, Bard, and DALL-E).

On behalf of the AIJA, I would like to thank the authors of the original study, Professor Lyria Bennett Moses (Faculty of Law and Justice UNSW), Dr Monika Zalnieriute (Faculty of Law and Justice UNSW), Professor Michael Legg (Faculty of Law and Justice UNSW), Dr Felicity Bell (then Faculty of Law and Justice UNSW) and Jake Silove (Senior Lawyer, Australian Government Solicitor). I also acknowledge Shahzeb Mahmood for his excellent research assistance on the revised report.

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

In the future, the AIJA aims to provide updates to this content on a regular basis through a web-based resource.

The Honourable Justice Murray Aldridge
Federal Circuit and Family Court of Australia,
President, Australasian Institute of Judicial Administration Inc.

Preface

The Australasian Institute of Judicial Administration (AIJA) engaged researchers at the UNSW Faculty of 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 was revised in 2023 and will continue to be updated on an ongoing basis. 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 original 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 roundtable 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.

Contents

1 Introduction

Artificial intelligence (AI) systems pervade modern life and are already being used in courts and tribunals, 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 a range of jurisdictions. 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 courts and tribunals. Section 4 discusses how AI tools, when used in the courtroom, impact on the core judicial values of open justice, accountability, and independence 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, tribunals and judiciary.

1 For example, in the recent case of *Thaler v Commissioner of Patents* (2021) 160 IPR 72; [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 was unanimously overturned in April 2022 by the Full Court of the Federal Court of Australia in *Commissioner of Patents v Thaler* (2022) 289 FCR 45; [2022] FCAFC 62.

2 Common AI Terms and Tools

2.1 Artificial Intelligence (AI)

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 terms 'complementary' or 'augmented' 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 originally defined 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 noted that 'AI systems are designed to operate with varying levels of autonomy'.[6] Other bodies define AI differently. A recent Discussion Paper on *Safe and Responsible AI* in Australia, from the Commonwealth Department of Industry, Science and Resources, defines AI as "engineered system that generates predictive outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives or parameters without explicit programming", also noting that such systems are designed to operate with varying levels of automation.[7] 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,[8] 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 systems apply 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 such as so-called 'if…then…' rules or using a series of logical statements to create a 'rule-set'. The former 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. The latter, logic programming, allows computers to draw inferences from given facts and relations.

2 Toby Walsh et al, *The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing* (Report, July 2019) 14.

3 John McCarthy et al, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (31 August 1955) <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>.

4 Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (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; Roger Clarke, 'The Re-Conception of AI: Beyond Artificial, and Beyond Intelligence' (2023) 4(1) *IEEE Transactions on Technology and Society* 24, 29–30.

6 Organisation for Economic Co-operation and Development, *Recommendation of the Council on Artificial Intelligence*, Doc No C/MIN(2019)3/FINAL, Meeting of the Council at Ministerial Level, 22-23 May 2019, 3 <https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf>. The definition was recently amended to include "content" in recognition of the importance of generative AI.

7 Australian Government, Department of Industry, Science and Resources, *Safe and responsible AI in Australia* (Discussion Paper, June 2023) 5 <consult.industry.gov.au/supporting-responsible-ai> ('*Safe and responsible AI Discussion Paper*').

8 See, eg, *Proposal for a 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* COM (2021) 206 final (21 April 2021) ('*Proposed AI Act*') <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.

Figure 1: Example of decision tree

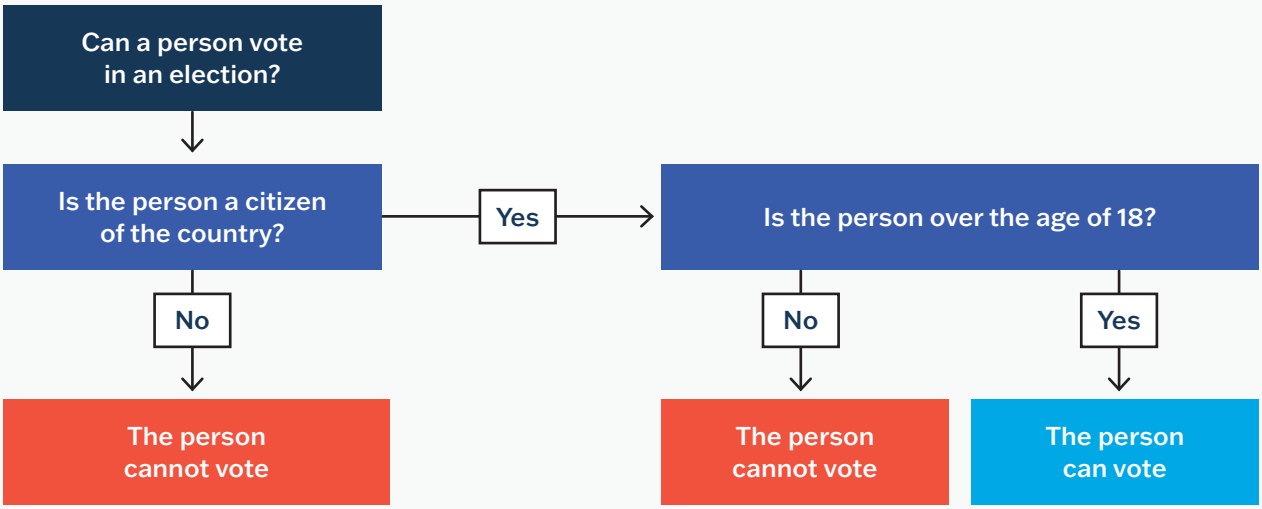


Figure 1 is an example of a 'binary' decision tree, as there are no more than two nodes stemming from each branch.

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, or a series of logic propositions.

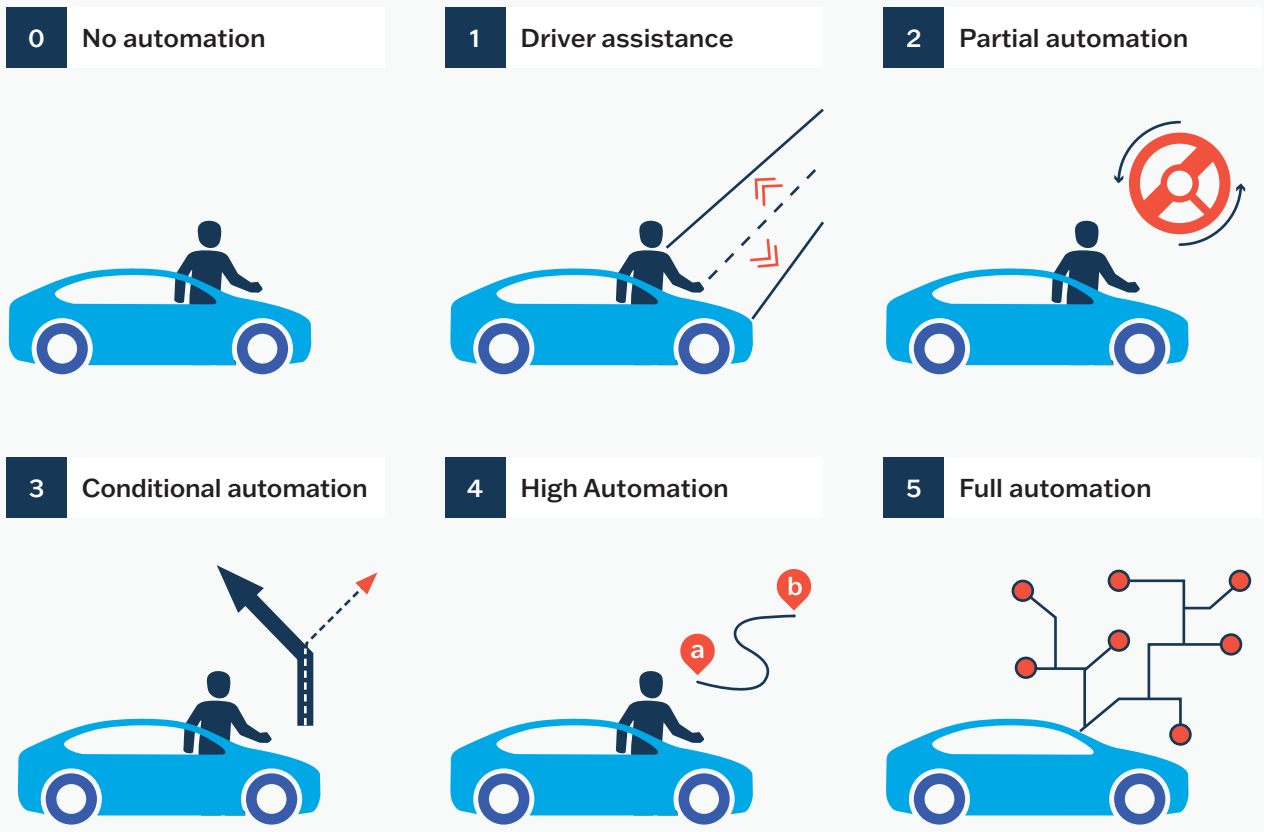
2.3 Automation

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]

An analogy can be drawn, albeit imperfectly, with automation in a court. At the lower end of the scale is a court 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 court 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 courts 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.

9 'SAE International Releases Updated Visual Chart for Its "Levels of Driving Automation" Standard for Self-Driving Vehicles', SAE International (Web Page, 2018) <https://www.sae.org/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>.

Figure 2: Levels of automation



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 2.7) 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.

10 Mireille Hildebrandt, 'Smart Technologies' (2020) 9(4) *Internet Policy Review 2*, 11 <https://policyreview.info/concepts/smart-technologies>.

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

2.4 Bot

The term bot refers to an 'agent' that acts autonomously usually in order to mimic a human. 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]

2.5 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.

2.6 Algorithm

The concept of an 'algorithm' pre-dates the first computer and was named after 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.

2.7 Machine Learning

Machine learning is 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 expert 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. When it is shown a cat of a different colour, for example, it may not recognise it as a cat because the system has already assumed that 'cat' is associated with the colour of cats in its training set. Human learners and computer 'learners' may thus be good at different tasks.

12 *Bolstering Online Transparency (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) 27-29 <https://www.digital.govt.nz/dmsdocument/95-better-rules-for-government-discovery-report>.

The use of machine learning for classification and clustering 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 (e.g. 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.

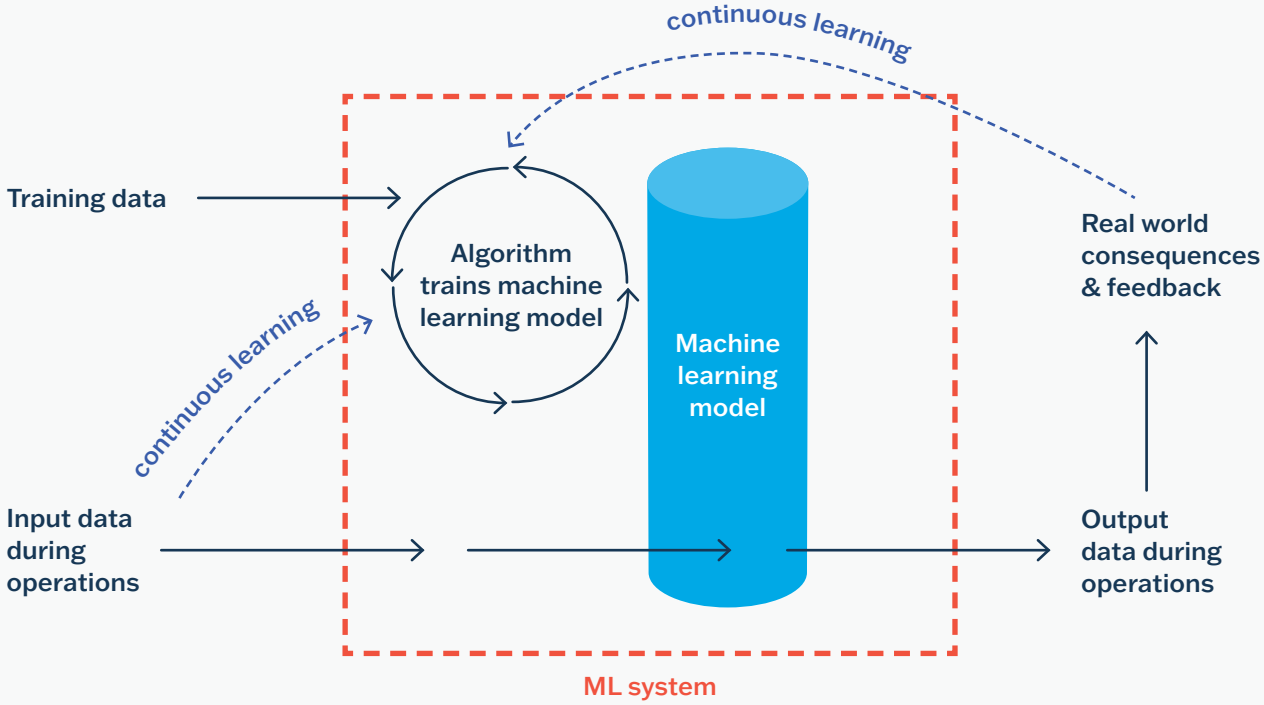
In *semi-supervised* machine learning, only some of the training data is labelled. These methods are often used where labelling data is expensive and time-consuming but unlabelled data is easy to obtain. Varying the above email clustering example, one might label a small number of emails in the training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

The more technical explanation in the remainder of this section 2.7 can be skipped for those less interested in technical detail.

Figure 3: Continuous learning



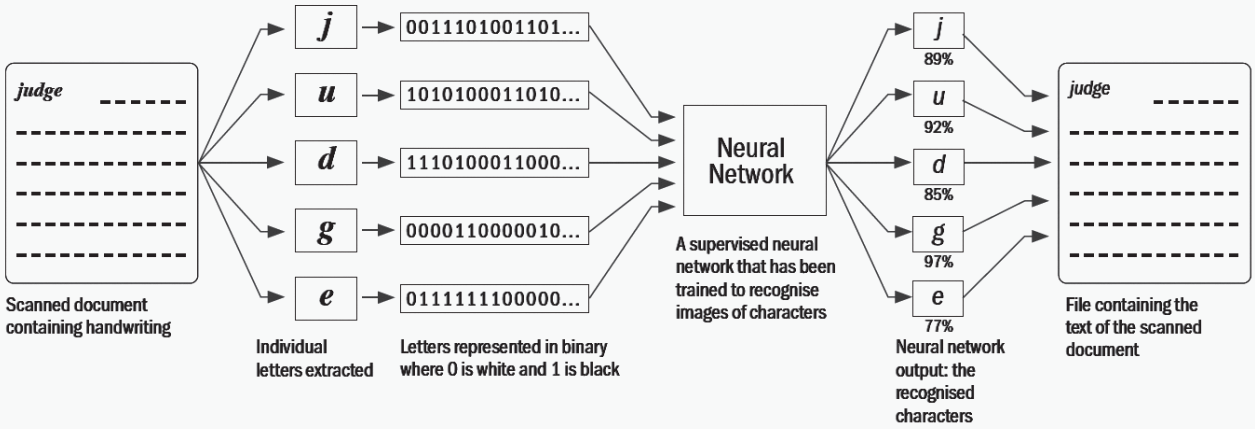
An example of a simple model is a *linear* model. This assumes a simple relationship between two variables (say x and y) where we assume that $y = mx + b$ (where m is the gradient of the line and b is the point of intercept with the y axis). In machine learning, an algorithm is used to train the model. 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 is only as good as the input data (see section 2.9 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 (including factoring in offsets and activation thresholds) 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).

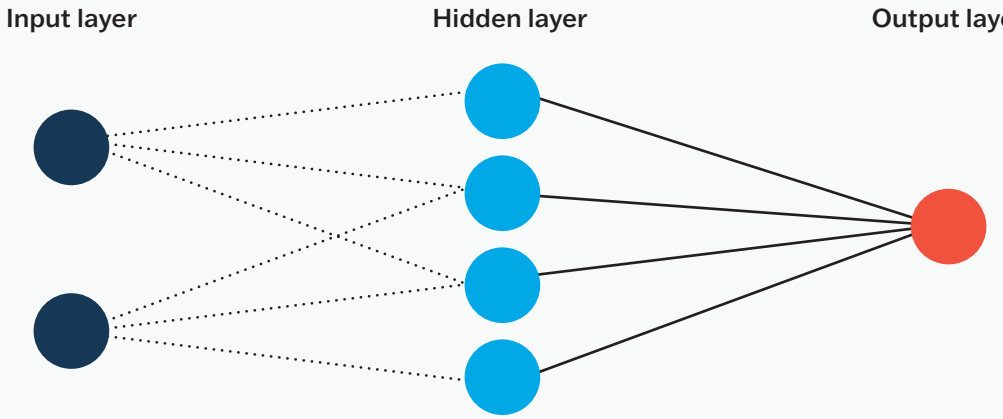
14 Theodore W. Ruger et al, 'The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking' (2004) 104(4) *Columbia Law Review* 1150, 1195-1205.

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 (with offsets and activation thresholds) 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).

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).

2.8 Generative AI

AI can be used for a variety of tasks. Thus far, most of the examples discussed involve tasks such as classification (e.g. a document is classified as discoverable or not discoverable, images are identified as cats or not cats) or clustering (e.g. emails are grouped into those with similar features).

Generative AI is used to describe an AI system capable of generating 'content', such as text, images or music, in response to prompts.[15] Examples include Google Translate, ChatGPT, Bard, and DALL-E. The most commonly discussed type of generative AI in the legal sphere are those generating text based on large language models (LLMs). These perform tasks such as answering questions and writing text in formats traditionally used by humans such as essays and poems. Generally applicable LLMs, such as GPT-4 (used in ChatGPT), can also be tuned to operate more effectively in specific contexts, such as law.[16] There is a technique in using these tools well, with the ability to write high quality prompts (associated with more useful outputs) an increasingly marketable skill.

Those interested in a more technical understanding of generative AI, and LLMs in particular, would need to look to:

1. **Neural network machine learning.** We have provided a high-level explanation at section 2.7 above. LLMs are trained on large volumes of text, for example taken from the Internet. Training a model may involve a combination of unsupervised, supervised, semi-supervised and reinforcement learning.[17]
2. **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'.[18] These techniques are used *inter alia* for translation, chatbots, search, and text generation.
3. **Transformers.** This is a model architecture that relies on an 'attention' mechanism, which are mathematical techniques to identify relationships between distant data elements (such as words in text). Without getting too deep into the technical mechanism, this architecture is good at inferring that 'it' refers to 'cat' in "The cat drank the milk because it was hungry" whereas 'it' refers to 'milk' in "The cat drank the milk because it was sweet".[19]
4. **Generative adversarial networks or GANs.** This involves a feedback process whereby systems provide feedback to each other. For example, one system generates a fake image of a person, whereas the other seeks to 'detect' whether the face is real. By running both learning processes in parallel, performance improves.

The ability to generate text, in particular, has led to questions about the usefulness of generative AI in producing legal texts (submissions, judgments, contracts, advice etc) as well as summaries and translations of such texts. Readers may wish to play around with generative AI tools to gain a sense of their capabilities. One can also find worked examples of what can be done in terms of, say, drafting documents in a (hypothetical) litigation online.[20]

15 Genevieve Bell et al, *Generative AI: Language models and multimodal foundation models* (Rapid Response Information Report, 24 March 2023) 2 <https://www.chiefscientist.gov.au/sites/default/files/2023-05/Rapid%20Response%20Information%20Report%20-%20Generative%20AI.pdf>

16 See 'Lexis +_ AI: Transform Your Legal Work', *LexisNexis* <https://www.lexisnexis.com/en-us/products/lexis-plus-ai.page>. However, the tool is currently focused on the US market.

17 Bell et al (n 15) 5.

18 Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart Publishing, 2020) 35.

19 Example from Ketan Doshi, 'Transformers Explained Visually (Part 1): Overview of Functionality', *Towards Data Science* (online, 14 December 2020) <https://towardsdatascience.com/transformers-explained-visually-part-1-overview-of-functionality-95a6dd460452>.

20 For an example in the context of litigation, see Kwan Yuen Iu and Vanessa Man-Yi Wong, 'ChatGPT by OpenAI: The End of Litigation Lawyers?' (2023) <https://tinyurl.com/2vt594hy>, although the tool it should be noted had access to the decided cases on which the facts were based.

Like other AI systems, generative AI systems based on LLMs have limitations. Some of these are discussed in the remainder of this section 2. One important point to bear in mind when using these systems is that the outputs are based on patterns in language and resulting predictions identify what word might come next in a particular context. If you ask ChatGPT to complete the sentence "The cat sat on the …", it will output "mat" not because it has observed cats sitting on mats more frequently than on chairs, but because it recognises the pattern and the fact that a sentence beginning in that way ends in "mat" more often than not. The outputs of generative AI systems might sometimes be true statements, but there is no guarantee that this will be the case based on how these systems function. In particular, there may be no 'truth filter' or source-checking, despite outputs that might suggest otherwise (e.g. "Yes, that is correct"). The term 'hallucinations' is sometimes used to describe outputs that suggest something is the case when it is not or where a non-existent source is cited. Those attributing sentience to tools such as ChatGPT fundamentally misunderstands its nature.[21] There are also significant risks in relying on LLMs.[22]

2.9 Garbage In – Garbage Out

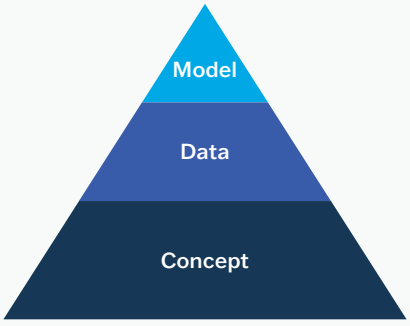
The reliance on data in machine learning means that the accuracy and reliability of the outputs generated will depend on the integrity and appropriateness of the training data that is used. If, for example, data collection was patchy so that it was systemically skewed, then the system will learn the 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.[23] Generative AI systems might similarly assume that a photo of "engineer" should be of a male given that the dataset from which a system is likely to be trained would reflect the currently male-dominated engineering field. This problem is often neatly summarised as 'garbage in – garbage out'. Conversely, training LLMs on reliable sources (such as textbooks) will increase the reliability of the model.[24]

2.10 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[25]



21 Bell et al (n 15) 2.

22 See, eg, Emily M. Bender et al, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (Conference Paper, Conference on Fairness, Accountability, and Transparency, March 3–10, 2021) <https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>.

23 Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women', *Reuters* (online, 11 October 2018) <https://www.reuters.com/article/us-amazon-comjobs-automation-insight-idUSKCN1MK08G>.

24 Suriya Gunasekar et al, Textbooks Are All You Need' (Cornell University Library, June 2023) 5.

25 Legg and Bell (n 18) 245.

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.[26]

2.11 Technological 'Black Box'

A technological 'black box' refers to a situation where the inner workings of some technological system are unknown or hidden. 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 does not 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.

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

2.12 Explainable AI

Explainable AI (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.[27] 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.[28]

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.[29] Similarly, a detailed explanation, constituting verification, will be required for a system determining the results of an election.[30] 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*.[31] 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?

27 Brent Mittelstadt, Chris Russell and Sandra Wachter, 'Explaining Explanations in AI' (Conference Paper, Conference on Fairness, Accountability, and Transparency, January 29–31, 2019) 279, 280.

28 'Welcome to the DataLex Community', *AustLII Communities* (Web Page, 24 August 2023) <http://austlii.community/foswiki/DataLex/WebHome>.

29 Lyria Bennett Moses and Edward Santow, 'Accountability in the Age of Artificial Intelligence: A Right to Reasons' (2020) 94(11) *Australian Law Journal* 829, 831

30 See Lyria Bennett Moses et al, 'No More Excuses: Automated Synthesis of Practical and Verifiable Vote-Counting Programs for Complex Voting Schemes' (Conference Paper, Electronic Voting: Second International Joint Conference, 2017) 66 <http://dx.doi.org/10.1007/978-3-319-68687-5_5>.

31 *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) ('GDPR').

3 Areas of AI Use in Courts

AI systems are increasingly used in litigation in jurisdictions around the world, ranging from Australia, China, the United States and the United Kingdom to India, 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".[32] 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 2.7 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.[33] 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'.[34]

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".[35]

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.[36]

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.[37]

32 Maura R Grossman and Gordon V Cormack, 'The Grossman-Cormack Glossary of Technology-Assisted Review' (2013) 7(1) *Federal Courts Law Review* 1, 32.

33 See, Supreme Court of Victoria, *Practice Note SC GEN 5: Technology in Civil Litigation* (First Revision, 29 June 2018) 8.9.

34 Legg and Bell (n 18) 112.

35 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* 221, 237.

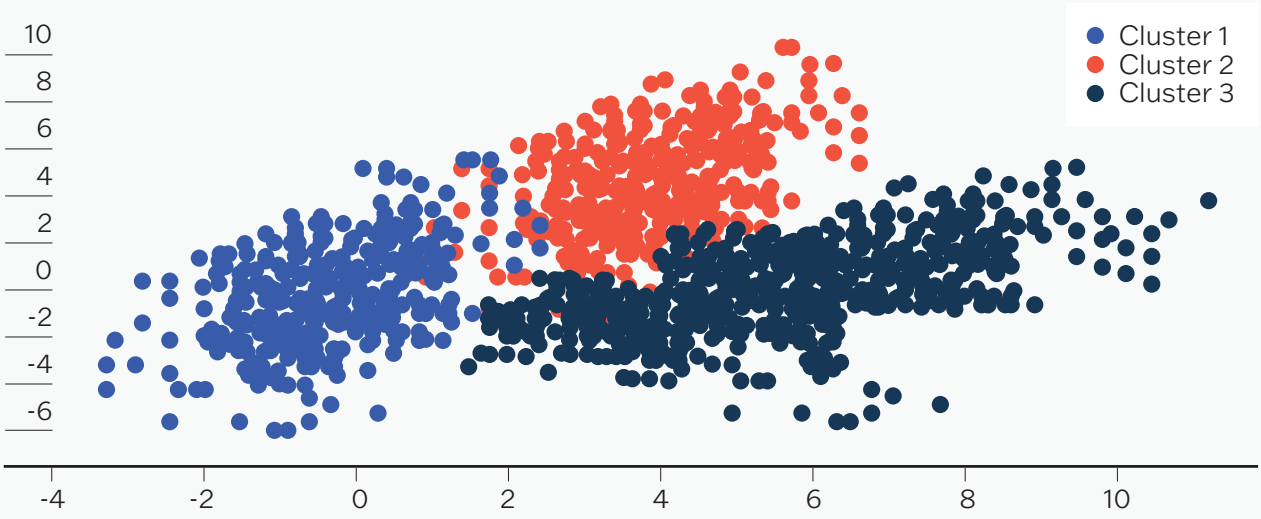
36 Grossman and Cormack (n 32). See also Jason Baron, Michael Berman and Ralph Losey (eds), *Perspectives on Predictive Coding and Other Advanced Search Methods for the Legal Practitioner* (American Bar Association, 2016).

37 Bolch Judicial Institute, Technology Assisted Review Guidelines (Guidelines, January 2019) 5; ViiV Healthcare Company v Gilead Sciences Pty Limited (No 2) [2020] FCA 1455, [135]-[138].

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

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.[38]

Figure 7: Example of clustering



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.[39] 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.[40]

Courts in the United States, Ireland, England and Wales, and Australia have approved the use of TAR in the litigation process.[41] 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).

38 Legg and Bell (n 18) 113-114.

39 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 'eDisclosure Systems Buyers Guide', *Complex Discovery* (Web Page, 2023) <https://complexdiscovery.com/buyers-guide/>.

40 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>

41 See *Da Silva Moore v Publicis Groupe* 287 FRD 182 (SDNY 2012) ('Da Silva Moore'); *Irish Bank Resolution Corporation Limited v Quinn* [2015] IEHC 175; *Pyrrho Investments Ltd v MWB Property Ltd* [2016] EWHC 256 (Ch) ('*Pyrrho Investments v MWB Property*'); *McConnell Dowell Constructors v Santam* [2016] VSC 734; *Parbery v QNI Metals Pty Ltd* [2018] QSC 276.

In *Pyrrho Investments Limited v MWB Property Ltd* [2016] EWHC 256 (Ch), the High Court of England and Wales endorsed the use of predictive coding software for e-disclosure, which reduced the documents subject to disclosure from over 17.6 million to 3.1 million. Master Matthews (as he then was) observed that algorithm-assisted document review was a useful tool in certain cases given the accuracy of predictive coding compared to manual review and keyword searches, the potential for greater consistency and cost-effectiveness, and the absence of restrictions in civil procedure rules.[42] Citing similar reasons, the High Court of England and Wales allowed the use of predictive coding system in *Brown v BCA Trading Ltd* [2016] EWHC 1464 (Ch), noting that the costs to use this tool was substantially less than that estimated for keyword searches.[43] More recently, Deputy High Court Judge David Halpern QC allowed predictive coding to determine relevance in cases involving a large volume of documents.[44] Ahead of the Rikki Neave trial at the Old Bailey, the criminal defence team used Luminance Discovery to analyse over 10,000 documents, saving £50,000 and a month of time of manual review.[45]

In *Da Silva Moore v Publicis Groupe* 287 FRD 182 (SDNY, 2012), United States Magistrate Judge Peck observed that the Court 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".[46] His Honour was acknowledging that the effectiveness of TAR could be assessed by reference to objective measures from the field of information retrieval, namely 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 because it demonstrates compliance with the orders for discovery to find and produce the relevant documents.[47] 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,[48] in choosing between types of TAR and in addressing disagreements as to statistical parameters.[49]

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.

42 Pyrrho Investments v MWB Property (n 41) [33] (Master Matthews).

43 Brown v BCA Trading Ltd & Ors [2016] EWHC 1464 (Ch) [3], [15] (Registrar Jones).

44 Isbilen v Turk & Ors [2022] EWHC 697 (Ch) [10] (Halpern J).

45 CRIMINAL COURT CASE DEFENCE PREPARATION' (Media Release, Luminance) <https://www.britishlegalitforum.com/wp-content/uploads/2022/01/Criminal_Court_Case_Defense_Preparation_2021.pdf>; Martin Bentham, 'Artificial Intelligence used for first time at Old Bailey by legal team in new Rikki Neave killing trial', *Evening Standard* (online, 3 December 2020) <https://www.standard.co.uk/news/crime/artificial-intelligence-rikki-neave-trial-old-bailey-b74380.html>.

46 *Da Silva Moore* (n 41) 184.

47 Legg and Bell (n 18) 114.

48 See, eg, *In re Mercedes-Benz Emissions Litigation* (D NJ, 2:16-cv-881 (KM) (ESK), 8 January 2020).

49 *McConnell Dowell Constructors (Aust) Pty Ltd v Santam Ltd (No 2)* [2017] VSC 640.

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.[50] 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.[51]

Alternative Dispute Resolution – ADR

Online Dispute Resolution – ODR – may be:

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

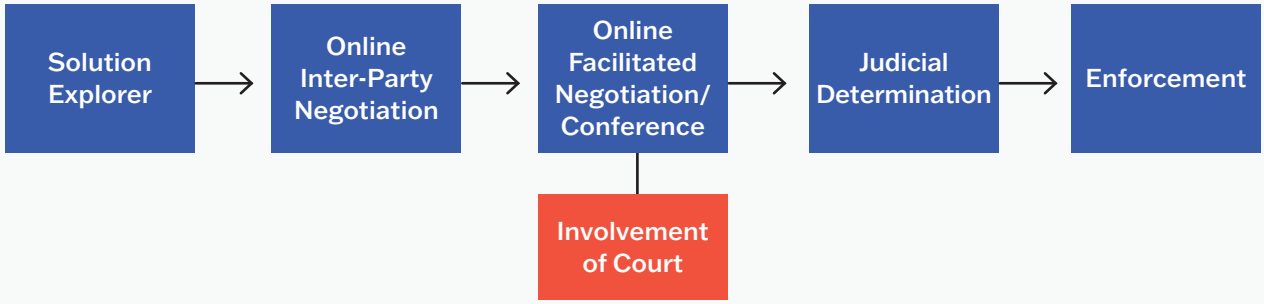
ODR may use a range of technologies, such as internet portals, email and audio-video conferencing facilities but, in relation to AI, ODR typically employs the expert system, or decision tree analysis (explained in section 2.2 Expert Systems and Traditional Programming). In the context of dispute resolution, 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:[52]

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.

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 8 shows.[53]

Figure 8: Integrated online court resolution of disputes



50 For example, eBay, PayPal and Alibaba.

51 Michael Legg, 'The Future of Dispute Resolution: Online ADR and Online Courts' (2016) 27 *Australasian Dispute Resolution Journal 227*, 227; Richard Susskind, *Online Courts and the Future of Justice* (Oxford University Press, 2019) 62–63.

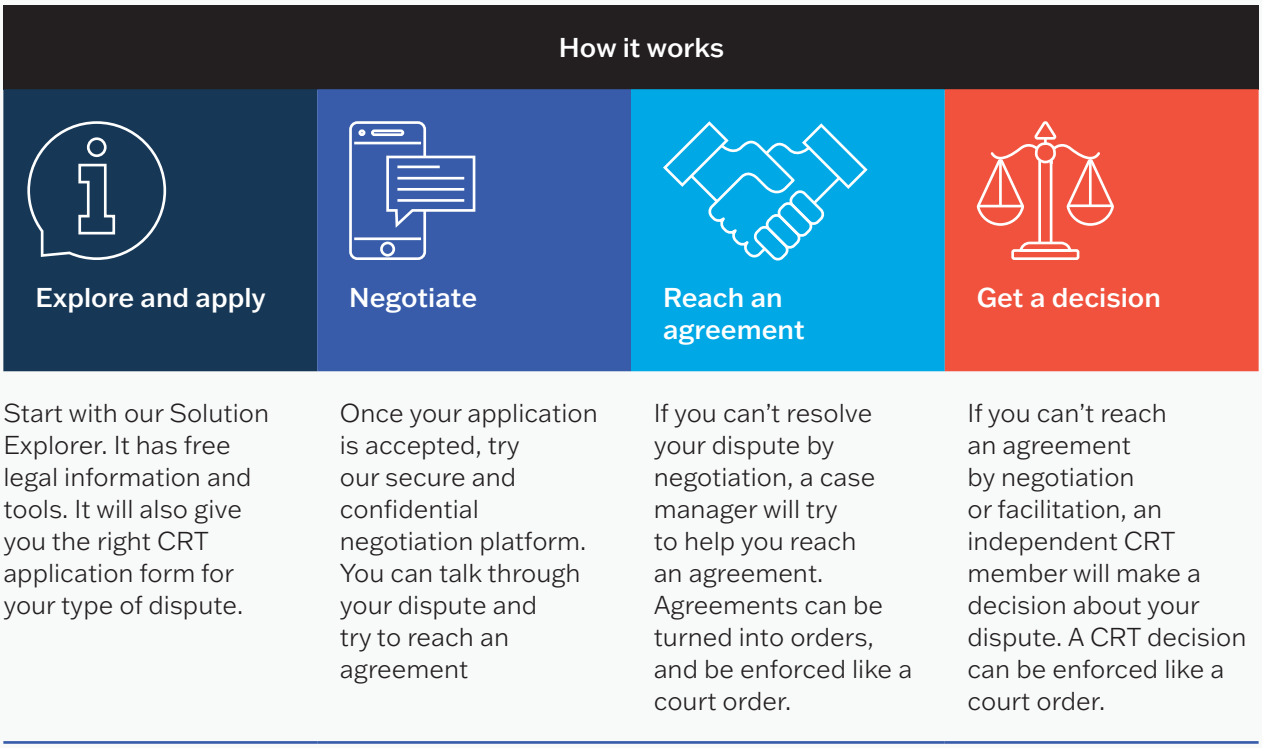
52 Legg and Bell (n 18) 138–9.

53 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.

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 (CRT).[54] 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.[55] The CRT is comprised of 21 tribunal members supported by a staff of 65 employees.[56] The CRT is composed of four steps as set out in Figure 9: The Civil Resolution Tribunal Process.

Figure 9: The Civil Resolution Tribunal Process



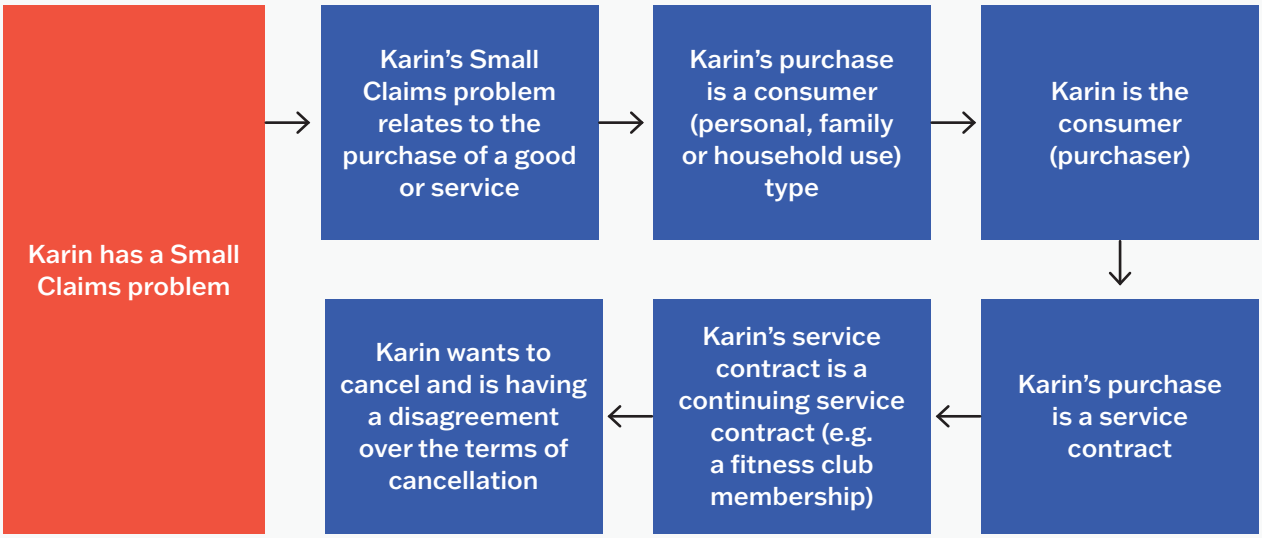
54 For other examples see Legg and Bell (n 18) 139-49.

55 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, 114; 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, 122.

56 *Civil Resolution Tribunal 2021/2022 Annual Report* (Report, 2022) 44-45 <https://civilresolutionbc.ca/wp-content/uploads/CRT-Annual-Report-2021-2022.pdf>.

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:

Figure 10: Example of the Solution Explorer expert system



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 and 37,903 times from 1 April 2021 to 31 March 2022. In 2021/2022, the average time to resolution for all dispute types was 92.7 days and the median time to resolution was 56 days for all dispute types.[57]

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.[58]

57 Ibid 32.

58 Traffic Penalty Tribunal, *Revolutionising a service* (Report, 2020) 13 <https://www.trafficpenaltytribunal.gov.uk/wp-content/uploads/2020/09/TPT_Revolutionising-a-Service_2020.pdf>.

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.[59] A comprehensive practice note, which supplements the *Civil Procedure Rules, Practice Direction 7E – Money Claim Online*,[60] 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).

A separate portal, made public in 2018 and known as the Civil Money Claims portal,[61] allows applicants to make a claim if the value of their loss is less than £25,000 (raised from £10,000 in May 2022).[62] Since its public beta testing in March 2018, more than 378,000 claims have made using the portal, with 97,315 claims filed in 2022 alone. It has settled 50.4% of the 9,560 mediation appointments made in 2022, within an average of 24 days. The system has achieved a 95% user satisfaction rating.[63]

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,[64] 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 2.12 Explainable AI).

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.[65] Machine learning relies on patterns in historic data. As mentioned in section 2.7 Machine Learning, a simple machine learning decision tree model could predict the outcome of US Supreme Court decisions.[66] 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,[67] 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.[68]

59 HM Courts & Tribunals Services, 'Money Claim Online (MCOL) – User Guide for Claimants' (2022) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/503066/money-claim-online-user-guide.pdf>.

60 'Practice Direction 7E – Money Claim Online', *Ministry of Justice* (Web Page, 10 August 2023) <https://www.justice.gov.uk/courts/procedure-rules/civil/rules/part07/pd_part07e>.

61 Letter from Lucy Frazer QC MP to Bob Neill MP, April 2018 <https://www.parliament.uk/globalassets/documents/commons-committees/Justice/correspondence/Lucy-Frazer-HMCTS-online-civil-claims-pilot.pdf> ('Letter from Lucy Frazer').

62 'Practice Direction 51R - Online Civil Money Claims Pilot', *Ministry of Justice* (Web Page, 10 August 2023) <https://www.justice.gov.uk/courts/procedure-rules/civil/rules/practice-direction-51r-online-court-pilot>; 'Make a Money Claim Online', *GOV.UK* (Web Page) <https://www.gov.uk/make-money-claim>.

63 'Fact sheet: Online Civil Money Claims', HM Courts and Tribunals Services (Web Page, 20 March 2023) <https://www.gov.uk/government/publications/hmcts-reform-civil-fact-sheets/fact-sheet-online-civil-money-claims#:~:text=achieved%20a%2095%25%20user%20satisfaction,days%20to%20reach%20a%20settlement>.

64 *Letter from Lucy Frazer* (n 61).

65 This seems to be the basis of the Singapore system. See 'Motor Accident Claims Online', Motor Accident Claims Online (Web Page) <https://motoraccidents.lawnet.sg/About>.

66 Ruger et al (n 14)

67 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', *ICPSR* (Web Page) <https://www.icpsr.umich.edu/web/ICPSR/series/86>. No equivalent database exists in Australia.

68 See Ashley, 'A Brief History of the Changing Roles of Case Prediction in AI and Law' (2019) 36(1) Law in Context A Socio-legal Journal 93.

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[69] and intellectual property lawsuits.[70]

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%.[71] Katz et al, who targeted their analysis at predicting the outcome of US Supreme Court decisions, were able to predict specific judicial votes in 240,000 instances with a 71.9% accuracy.[72] 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.[73] 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 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.[74] 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 also 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.[75]

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.[76] 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 a 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.[77] 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.[78]

69 See 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.

70 Mihai Surdeanu et al, 'Risk Analysis for Intellectual Property Litigation' in *International Conference on Artificial Intelligence and Law* (2011) 116, 116–120.

71 Octavia-Maria Sulea et al, 'Predicting the Law Area and Decisions of French Supreme Court Cases' [2017] *arXiv:1708.01681 [cs]* 5 <http://arxiv.org/abs/1708.01681>.

72 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 1, 8.

73 See Nikolaos Aletras et al, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 *PeerJ Computer Science* e93; 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.

74 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.

75 Frank Pasquale and Glyn Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism' (2018) 68(1) University of Toronto Law Journal 63, 79.

76 Medvedeva, Vols and Wieling (n 73) 256, 259-262.

77 Brian M Barry, *How Judges Judge: Empirical Insights into Judicial Decision-Making* (Informa Law from Routledge, 2021) 278.

78 Meng Yu and Guodong Du, 'Why Are Chinese Courts Turning to AI?', *The Diplomat* (online, 19 January 2019) <https://thediplomat.com/2019/01/why-are-chinese-courts-turning-to-ai/>; Sourdin (n 11) 14.

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 have analysed migration decisions, though emphasising that their goals are descriptive rather than necessarily predictive.[79]

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.[80] There are also tools used for specific purposes, such as litigation funding.[81]

Two of the most high-profile examples of litigation analytics are Lex Machina (developed by the IP Litigation Clearinghouse at Stanford University) and Context (previously called Ravel) both of which were purchased by LexisNexis. Both chiefly utilise US data and apply to US courts. LexMachina works by obtaining court and patent office data every 24 hours which it codes using a proprietary NLP and ML engine called Lexpressions™. For every case, LexMachina extracts the patent involved, the players (lawyers, judges, parties) and legal data such as findings and outcomes, including any damages awarded. The variables that are extracted are analysed to look for 'meaningful patterns' that provide insights into how a future case may resolve. The AI tool behind LexMachina has been applied to other litigation areas such as competition law, employment law and insurance. Ravel made a 'Judge Analytics' tool available in 2015. The tool claims to 'surface' the most persuasive language to use depending on the judge and court. It combines NLP and ML, but also uses design principles to better communicate its insights. The underlying technology has also been used by LexisNexis to also provide insights as to opposing lawyers and expert witnesses.[82]

These tools purport to predict how judges will decide particular motions (e.g. summary judgment, admission of evidence) and even the outcome of entire cases, so as to assist lawyers to optimise the precedents they rely on and the arguments they make. The insights from the tools are in addition to, rather than a replacement for, traditional legal research.

The use of statistics and prediction is also relevant in the context of criminal proceedings, and in particular to predict sentencing outcomes.[83] 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.

79 'Study Raises Questions about Visa Appeals', *AMES Australia* (online, 12 March 2020) <https://amesnews.com.au/latest-articles/study-raises-questions-about-visa-appeals/>; Hagar Cohen, 'Almost 99 per cent of protection visa review applications fail when heard by controversial 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>.

80 'Legal Analytics – Quickly Uncover Strategic Information', *Lex Machina* (Web Page) <https://lexmachina.com/legal-analytics/>.

81 'Litigation Finance', *Legalist* (Web Page) <https://www.legalist.com/strategies/litigation-finance>.

82 Legg and Bell (n 18) 92-93. Further information is available at Lex Machina (Web Page) <https://lexmachina.com/> and 'Predictive Data Analytics for Legal Language', LexisNexis (Web Page) <https://www.lexisnexis.com/en-us/products/context.page>. Other examples include Blue J legal which predicts judicial rulings on Canadian tax laws and eLegPredict which aims to predict decisions of the Indian Supreme Court.

83 See generally 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[84] 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.[85] For example, the Correctional Offender Management Profiling for Alternative Sanctions tool (COMPAS)[86] 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".[87] 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).[88] 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; by 2018, COMPAS had been used to assess over one million offenders.[89] 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).[90]

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.[91]

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

85 Carolyn McKay, 'Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making' (2019) 32(1) *Current Issues in Criminal Justice* 22, 31.

86 Developed by Equivant (previously Northpointe). See *Equivant* (Web Page) <https://www.equivant.com/>.

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

88 Equivant, 'Practitioner's Guide to COMPAS Core' (4 April 2019) 31-32 <https://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf>.

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

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

91 Alyssa M Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103 *Iowa Law Review* 303, 319; Derek Thompson, 'Should We Be Afraid of AI in the Criminal-Justice System?', *The Atlantic* (online, 20 June 2019) <https://www.theatlantic.com/ideas/archive/2019/06/should-we-be-afraid-of-ai-in-the-criminaljustice-system/592084/>; Julia Angwin et al, 'Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks.', *ProPublica* (online, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

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.[92] 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:[93]

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.[94] However, defence counsel has no correlated right to challenge the accuracy or methods of the COMPAS program,[95] and both the defendant and the judge may have limited information about the reliability of the tool.[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'.[96] 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.[97] In 2017 Loomis' petition to the US Supreme Court was denied.[98]

There has been some research in this space, both on the issue of predictive accuracy as well as on the issue of bias. The results in respect of the former are mixed. In various studies, predictive accuracy has been found: to have a similar reliability rate as individuals with no expertise given only two variables to consider; to be less reliable at predicting violent crime compared to criminal behaviour more broadly; and to have good predictive accuracy at the extreme ends (those with the highest risk score reoffended at nearly four times the rate of those with the lowest risk score).[99] Even where studies have been conducted, one needs to be careful in drawing conclusions – those studies all inevitably use a proxy for reoffending (such as arrest) that may not indicate actual reoffending.

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

93 Ibid [66].

94 Ibid [56].

95 Ibid [51]; McKay (n 85) 11.

96 *Loomis* (n 92) [86].

97 Ibid [98].

98 *Loomis v Wisconsin*, (No 16-6387, 26 June 2017).

99 Florence G'sell, 'AI Judges' in Larry A. DiMatteo, Cristina Poncibò and Michel Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge University Press, 2022) 347, 350.

In addition to questions about predictive accuracy, there are very serious issues with bias. 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).[100] This was cited in *Michigan v Canedo*, although the use of COMPAS was not an issue fully raised by the defendant in that case.[101]

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.

The use of criminal sentencing and risk assessment tools extends beyond the United States and has generated mixed responses. For example:

- Singapore's Chief Justice Sundaresh Menon indicated in October 2022 that the likelihood of Singapore adopting AI tools for criminal sentencing in the near future was low.[102]
- The Office of Chief Registrar of the Federal Court of Malaysia announced plans for phased implementation of AI-guided sentencing guidelines in the sessions and magistrate courts.[103] In February 2020, the judiciary in Sabah and Sarawak in Malaysia was the first in the country to use AI in sentencing. In Public *Prosecutor v Denis P. Modili* [2020] 2 SMC 381,[104] the Malaysian magistrates court used a sentence recommendation tool as a "mere guideline to assist the Court so as not to depart from the true spirit of a reasonable sentencing principle".[105] However, lawyers in Malaysia and the Malaysian Bar Council have expressed reservation about the move to implement predictive algorithms for criminal sentencing.[106]
- Taiwan introduced an AI tool to assist citizen judges in sentencing decisions. This tool allows citizen judges to input case-related information and, based on analyses of the factual circumstance (such as any admission of guilt and mitigating circumstances), the sentencing guidelines and past decisions by professional judges, makes a range of recommendations.[107]

100 Angwin et al (n 91); 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 also Andrew Lee Park, 'Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing', *UCLA Law Review* (online, 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.

101 *People v. Canedo* 961 N.W.2d 763 (Mich Sup Ct, 2021).

102 CNA, 'No plans for Singapore's criminal courts to use AI in sentencing for now' (YouTube, 1 November 2022) <https://www.youtube.com/watch?v=IJwbnn2v2_A&ab_channel=CNA>; Selina Lum, 'S'pore not likely to use AI in sentencing in foreseeable future: Chief Justice', *The Strait Times* (online, 31 October 2022) <https://www.straitstimes.com/singapore/courts-crime/s-pore-not-likely-to-use-ai-in-sentencing-in-foreseeable-future-chief-justice>.

103 V Anbalagan, 'Malaysian Bar troubled over judges using AI for sentencing', *Free Malaysia Today* (online, 24 July 2021) <https://www.freemalaysiatoday.com/category/nation/2021/07/24/malaysian-bar-troubled-over-judges-using-ai-for-sentencing/>.

104 *PP v Denis P Modili* [2020] 2 SMC 381 (Magistrate Court, Kota Kinabalu).

105 See generally Dennis W K Khong and Chiung Ching Ho, 'Case Commentary: Artificial Intelligence in Malaysian Courts: PP v Denis P Modili' (2022) 2(2) *Asian Journal of Law and Policy* 127. See also 'ARTIFICIAL INTELLIGENCE (AI)', *e-Kehakiman Sabah and Sarawak* (Web Page) <https://ekss-portal.kehakiman.gov.my/portals/web/home/article_view/0/5/1>.

106 Danial Martinus, 'Malaysia tests AI court sentencing despite ethical concerns raised by lawyers: The AI sentenced two men in Sabah in 2020', Mashable SE Asia (online, 13 April 2022) <https://sea.mashable.com/tech/20026/malaysia-tests-ai-court-sentencing-despite-ethical-concerns-raised-by-lawyers>; Anbalagan (n 103).

107 TaiwanPlus News, 'Taiwan Launches AI Sentencing Tool for Citizen Judges | TaiwanPlus News' (YouTube, 7 February 2023) <https://www.youtube.com/watch?v=HmUwcL1ZbR4&ab_channel=TaiwanPlusNews>.

- China has embraced sentence recommendation systems. In Shanghai, Hainan and Guangzhou, the sentence recommendation system enables judges to examine historical cases with comparable factual patterns and presents the judge with a sentence recommendation.[108] Leveraging big data and AI technologies, the sentence standardisation process seeks to ensure that key information is accurately extracted and considered.[109] Where a judge deviates from the recommendation, they must provide written reasons.[110] Another system, the Xiao Baogong Intelligent Sentencing Prediction System,[111] allows judges and prosecutors in China to use big data analysis to extract information about defendants, charges, conviction conditions, aggravating and mitigating circumstances, and regions to recommend penalties in real time. The system provides judges with ten contextually relevant cases within the region for each prediction.[112]

While there has been less scrutiny and judicial consideration of these criminal sentencing and risk assessment tools, the concerns raised in respect of COMPAS are very likely to apply equally to any tool that seeks to predict and make recommendations to judges in the context of matters which do or may affect the liberty of individuals. 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.*[113]

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,[114] 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.

In their submissions to the Human Rights Council Advisory Committee in October 2022, Zalnieriute and Cutts called for a ban on the use of predictive algorithms in criminal sentencing, noting that such tools not only "impose burdens upon people of colour without justification; they also reinforce unjust views that are liable to result in the systemic denial of opportunities to people of colour across many aspects of public and private life".[115]

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. A variety of views have been expressed on the appropriateness of using these tools.

108 Rachel E. Stern et al, 'Automating Fairness? Artificial Intelligence in the Chinese Court' (2021) 59 *Columbia Journal of Transnational Law* 515, 526-527.

109 Yuan Shenggao, 'AI-assisted sentencing speeds up cases in judicial system', *China Daily* (online, 18 April 2019) <https://www.chinadaily.com.cn/cndy/2019-04/18/content_37459601.htm>.

110 Ben Wodecki, 'AI helps judges decide court cases in China', *AI Business* (online, 19 July 2022) <https://aibusiness.com/verticals/ai-helps-judges-decide-court-cases-in-china>.

111 *Xiao Bao Gong Law AI* (Web Page) <https://www.xiaobaogong.com/>.

112 Alena Zhabina, 'How China's AI is automating the legal system', *DW* (online, 20 January 2023) <https://www.dw.com/en/how-chinas-ai-is-automating-the-legal-system/a-64465988#:~:text=In%20China%2C%20people%20can%20use,can%20even%20calculate%20legal%20costs>.

113 *Wong v The Queen* (2001) 207 CLR 584 [59]

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

115 Monika Zalnieriute and Tatiana Cutts, 'How AI and New Technologies Reinforce Systemic Racism', Submission to the United Nations Human Rights Council, 3 October 2022, 4 <https://www.ohchr.org/sites/default/files/documents/hrbodies/hrcouncil/advisorycommittee/study-advancement-racial-justice/2022-10-26/HRC-Adv-comm-Racial-Justice-zalnieriute-cutts.pdf>.

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.[116]

Another example in the judicial context is the EXPERTIUS system in Mexico,[117] which advises 'novice' judges and clerks as to whether a plaintiff is eligible for a pension in addition to the quantum of that pension.[118] 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).[119]

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.[120] 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.[121] 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".[122] 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 service that would be "just, proportionate, accessible to all and works better for everyone".[123] The proposal stalled after the 2017 general election and does not seem to have been revived.[124]

116 Chief Justice TF Bathurst, 'Modern and Future Judging' (Speech, Sir Maurice Byers Lecture 2021, 3 November 2021) [48].

117 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.

118 Ibid 78; Davide Carneiro et al, 'Online Dispute Resolution: An Artificial Intelligence Perspective' (2014) 41 *Artificial Intelligence Review* 211, 227–228.

119 Carneiro et al (n 118) 227–228.

120 Joshua Rozenberg QC, 'Automatic online conviction', *The Legal Education Foundation* (Web Page, July 2020) <https://long-reads.thelegaleducationfoundation.org/automatic-online-conviction/>; Owen Bowcott, 'Government's £1bn Plan for Online Courts 'Challenges Open Justice'', *The Guardian* (online, 15 March 2017) <http://www.theguardian.com/law/2017/mar/15/governments-1bn-plan-for-online-courts-challenges-open-justice>.

121 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 ('*Government Response No Cm 9391*'). 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.

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

123 *Government Response No Cm 9391* (n 121) [21(b)].

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

In Brazil, at least 47 courts have AI programs and systems in use or under development.[125] The Brazil Supreme Federal Tribunal uses a software that automates examination of appeals and provides recommendations on, among other things, legal precedents and potential courses of action. Similarly, the Tribunal of Justice of Minas Gerais employs software capable of identifying and categorising legal resources that relate to the same issues or are the subject of applicable precedents.[126]

The Supreme Court of India launched the Supreme Court Portal for Assistance in Courts Efficiency (SUPACE) in 2020, which assists judges with accessing information.[127] This machine learning application offers diverse features, including file previews, a chatbot for case overview and queries, a universal search function, real-time progress tracking, work detail information, a logic gate for fact extraction, and a notebook for preparing summary documents.[128]

In March 2022, Saudi Arabia introduced virtual enforcement courts that operate without human intervention, which is said to streamline a previously 12-step litigation process down to two steps.[129] In the United Arab Emirates, the Abu Dhabi Judicial Department introduced a smart court initiative in August 2022, leveraging AI to enhance adjudication rate and expedite adjudication process.[130]

In 2019, reports have emerged that the Estonian Ministry of Justice has sought to automate the adjudication of small contract disputes.[131] Many articles in popular technology magazines claimed that 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 was meant to "adjudicate small claims disputes of less than €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'.[132] 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.[133] Notwithstanding the interest surrounding these purported announcements, Estonia's government later clarified that the reports were misleading and that there were no plans for automated courts to be implemented in Estonia.[134]

China has perhaps gone the furthest in creating 'smart courts' that resolve disputes through an on-line platform. According to the Supreme People's Court, a sophisticated system will be established to support integration of AI in the judicial sector by 2025.[135] Chinese authorities have already established internet courts in Beijing, Guangzhou, and Hangzhou. These so-called 'smart courts' are said to offer a convenient alternative to traditional litigation by allowing participants to register cases, serve documents, present evidence and

125 Eduardo Villa Coimbra Campos, 'Artificial Intelligence, the Brazilian Judiciary and Some Conundrums', *SciencesPo* (Blog Post, 3 March 2023) <https://www.sciencespo.fr/public/chaire-numerique/en/2023/03/03/article-artificial-intelligence-the-brazilian-judiciary-and-some-conundrums/>.

126 Katie Brehm et al, *The Future of AI in Brazilian Judicial System: AI Mapping, Integration, and Governance* (Report, 2020) 14 <https://itsrio.org/wp-content/uploads/2020/06/SIPA-Capstone-The-Future-of-AI-in-the-Brazilian-Judicial-System-1.pdf>.

127 Aamir Khan, 'AI-powered Indian judiciary: A step forward or cause for concern?', *Bar and Bench* (Blog Post, 6 June 2023) <https://www.barandbench.com/columns/litigation-columns/ai-powered-indian-judiciary-a-step-forward-cause-concern>.

128 Samiksha Mehra, 'AI is set to reform justice delivery in India', *INDIAai* (Blog Post, 6 April 2021) <https://indiaai.gov.in/article/ai-is-set-to-reform-justice-delivery-in-india>.

129 'Justice minister inaugurates Virtual Enforcement Court in Saudi Arabia', *Zawya* (online, 28 March 2022) <https://www.zawya.com/en/legal/justice-minister-inaugurates-virtual-enforcement-court-in-saudi-arabia-p5dhnpu7>.

130 Abdulla Rasheed, 'Abu Dhabi criminal cases now followed up by artificial intelligence', *Gulf News* (online, 8 August 2022) <https://gulfnews.com/uae/crime/abu-dhabi-criminal-cases-now-followed-up-by-artificial-intelligence-1.89792712>.

131 Victor Tangermann, 'Estonia Is Building a "Robot Judge" to Help Clear Legal Backlog', *Futurism* (online, 25 March 2019) <https://futurism.com/the-byte/estonia-robotjudge>; Tara Vasdani, 'Estonia set to introduce 'AI judge' in small claims court to clear court backlog', LAW360 Canada (online, 10 April 2019) <https://www.law360.ca/articles/11582/estonia-set-to-introduce-ai-judge-in-small-claims-court-to-clear-court-backlog->.

132 Eric Niler, 'Can AI Be a Fair Judge in Court? Estonia Thinks So', WIRED (online, 25 March 2019) <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/>.

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

134 'Estonia does not develop AI Judge', *Republic of Estonia Ministry of Justice* (Web Page, 16 February 2022) <https://www.just.ee/en/news/estonia-does-not-develop-ai-judge>.

135 'The Supreme People's Court The Opinions on Regulating and Strengthening the Applications of Artificial Intelligence in the Judicial Fields (2022)', *China Laws Portal – CJO* (Web Page, 8 December 2022) <https://www.chinajusticeobserver.com/law/x/the-supreme-people-s-court-the-opinions-on-regulating-and-strengthening-the-applications-of-artificial-intelligence-in-the-judicial-field-20221208>. See also Cao Yin, 'AI assistance to boost efficiency of judicial sector', *China Daily* (online, 12 December 2022) <https://www.chinadaily.com.cn/a/202212/12/WS6396843da31057c47eba3e3c.html>.

resolve disputes using the online platform without having to appear in person. These courts handle a variety of disputes, including intellectual property, e-commerce, financial disputes related to online activities, domain name issues, product liability arising from online purchases, and certain administrative disputes. Between September 2018 and August 2019, the Beijing Internet Court handled over 26,000 copyright ownership and infringement disputes and over 4,000 online shopping contract disputes, accounting for 77.7% and 12.3% of disputes in these categories, respectively.[136] Overall, the court accepted a total of 34,263 cases and finalised 25,333 cases in this timeline. Additionally, between January 2019 to May 2023, the court also handled 272 cases involving minors, with minors predominantly serving as plaintiffs. These cases were primarily resolved through either mediation or withdrawal.[137] The court has introduced other digital innovations, such as identity authentication through facial recognition, automatic generation of pleadings, litigation risk assessment, real-time voice recognition for court records, online electronic signatures, document writing assistance, one-click document delivery, and an offline panoramic experience to educate the public on online litigation rules and operations.[138]

3.6 Automated E-Filing

Electronic filing (e-filing) of documents in court/tribunal proceedings has become ubiquitous in modern court systems. In October 2022, for instance, China's Supreme People's Court reported that 11,439,000 cases were filed electronically in 2021.[139] 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.[140]

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.[141]

Automated e-filing can extend into automated verification. Starting from March 2023, an internet-based judicial blockchain platform has been implemented across more than 3,500 courts in China to enable online verification of electronically served documents.[142] This platform allows parties or trusted third parties to access and verify the authenticity of documents received electronically from the courts. For example, where a bank receives an electronically served document that requires the release of a debtor's deposits to a creditor, the bank can use the platform to verify the document's authenticity.

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).

136 *White Paper on Trials of Beijing Internet Court* (White Paper, 14 October 2019) 5 <https://regional.chinadaily.com.cn/pdf/WhitepaperontrialsofBeijingInternetCourt.pdf> ('*Beijing Internet Court White Paper*').

137 *White Paper on Judicial Protection of Minors on the Internet* (White Paper, 4 July 2023) <https://www.chinadaily.com.cn/specials/BeijingInternetCourtWhitePaperonJudicialProtectionofMinorsontheInternet.doc>.

138 *Beijing Internet Court White Paper* (n 131) 14, 17, 19, 21, 22, 38.

139 '10 Million E-filings in Chinese Courts in 2021', *China Justice Observer* (Web Page, 28 November 2022) <https://www.chinajusticeobserver.com/a/10-million-e-filings-in-chinese-courts-in-2021>.

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

141 Ibid.

142 'E-Documents From Chinese Courts Can Be Verified on Blockchain', *China Justice Observer* (Web Page, 8 May 2023) <https://www.chinajusticeobserver.com/a/e-documents-from-chinese-courts-can-be-verified-on-blockchain>.

THINGS TO CONSIDER – Questions for courts and tribunals

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. From 2019, CE-File was made mandatory for legally represented parties in the Business and Property Courts throughout the UK.
- 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. The NextGen CM/ECF system is similar to CE-File and acts as a comprehensive case management system. PACER gives the public access to over 1 billion documents filed in more than 200 federal courts.
- In Australia, the National Court Framework, adopted by the Australian Federal Court in 2014, streamlined and synchronised the operation of State registries and the operation of individual judges' dockets. E-filing is now offered by most Australian courts.

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. 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). 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. 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, and consequently e-filing is available continuously and not only when the court is open. 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.

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.

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",[143] 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.[144]

143 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(3) *Hastings Law Journal* 589, 597.

144 Ibid 598–601.

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.[145] In 2017 alone, 80.1 million matters were awaiting judicial determination in Brazil, many of which have been described as "routine and low value".[146] 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.[147] 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.[148] The ATHOS system, used by the Superior Court of Justice in Brazil, is similar. This system focuses on 'repetitive appeals', identifying cases that can be treated collectively.[149]

Similarly, AI systems can be used to direct the attention of the court.[150] 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?".[151] 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.[152] 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.[153] 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.

145 'Inteligência Artificial: Trabalho Judicial de 40 Minutos Pode Ser Feito Em 5 Segundos', *Supremo Tribunal Federal* (Web Page, 23 October 2023) <http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=393522>.

146 Susskind (n 51) 290

147 Pedro HG Inazawa et al, 'Project Victor' (2019) 39(1) *Revista Computacao Brasil, Sociedade Brasileira de Computacao* 19, 20.

148 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.

149 Luisa Hedler, 'Time, Law, and Tech: The Introduction of Algorithms to Courts of Law' (PhD Series No. 17.2023, Copenhagen Business School, 2023), 74-76.

150 Ryan W Copus, 'Statistical Precedent: Allocating Judicial Attention' (2020) 73 *Vanderbilt Law Review* 605.

151 Ibid 611.

152 [2018] WASC 310.

153 Ibid.

3.8 Natural Language Processing and Generative AI

In Australia, services such as Auscript, Transcription Australia and Epiq provide courts with transcription services, some of which boast real-time transcription. For instance, EpiqFAST is specifically designed for courtrooms. Unlike a dictation-type environment, where only one person is speaking and so a speech recognition engine is trained to get a good output from one voice, this fully automated speech-to-text program can recognise different people speaking into different microphones and generate real-time transcripts with 98% accuracy rate.[154] IBM has achieved a 5.5% word error rate (compared to the standard human error rate of 5.1%),[155] with a "dramatic improvement in accuracy" driving the likelihood that court reporting will increasingly be an automated process.[156] In July 2022, VIQ Solutions commenced its multi-year contract with the Queensland Courts Department of Justice and Attorney General to provide courtroom monitoring and transcription services for approximately 50% of court cases in Queensland.[157] Voice recognition and transcription can be automated, and globally the speech recognition market is expected to be worth at least USD 18 billion by 2023.

The New Zealand Office of the Chief Justice issued the Digital Strategy for Courts and Tribunals issued in March 2023, outlining a range of initiatives to be pursued by New Zealand in the next decade. These initiatives include implementing automated speech-to-text transcription services for hearings and integrating automated interpretation service.[158]

Some Chinese courts use real-time voice recognition to produce court transcripts.[159] iFLYTEK is a technology company used during some trials which translate real-time audio into Mandarin and English text.[160] In the Liaoning Higher People's Court, an intelligent voice technology has been integrated to transcribe court hearings.[161] Over 4,200 courtrooms across 31 out of China's 34 provincial administrative regions, including Shanghai,[162] has adopted an AI-supported software called the 206 System[163] that can accurately identify speakers based on their roles (such as judges, prosecutors and defendants) and transcribe speech in real-time. The software reportedly improves transcription speed for law clerks from 120-150 words to 250-350 words per minute, reducing overall trial time by approximately 30%.[164]

154 Ksenia Stepanova, 'The technology that's transforming court reporting', *Australian Lawyer* (online, 4 April 2023) <https://www.thelawyermag.com/au/news/general/the-technology-thats-transforming-court-reporting/441199>.

155 'What Is Speech Recognition?', *IBM* (Web Page) <https://www.ibm.com/cloud/learn/speech-recognition>.

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

157 'VIQ Solutions Commences Multi-Year Contract with Queensland Courts Department of Justice and Attorney General' (Press Release, VIQ Solutions, 27 June 2022) <https://viqsolutions.com/media-center/viq-commences-multi-year-contract-with-qld-courts-djag-attorney-general/>.

158 The Office of the Chief Justice, *Digital Strategy for Courts and Tribunals* (Report, March 2023) 26 <https://www.courtsofnz.govt.nz/assets/7-Publications/2-Reports/20230329-Digital-Strategy-Report.pdf>.

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

160 Lin Haibin, Guodong Du and Meng Yu, 'Big Data, AI and China's Justice: Here's What's Happening', *China Justice Observer* (online, 1 December 2019) <https://www. chinajusticeobserver.com/a/big-data-ai-and-chinas-justice-heres-whats-happening>; Mara Hvinstendahl, 'How a Chinese AI Giant Made Chatting—and Surveillance—Easy', *WIRED* (online, 18 May 2020) <https://www.wired.com/story/iflytek-china-ai-giant-voice-chatting-surveillance/>.

161 '全省128家法院实现智能语音应用全覆盖，创造辽宁"智慧法院"新高度' [128 courts in the province have achieved full coverage of intelligent voice applications, creating a new level of 'smart courts' in Liaoning] Liaoning Smart Court (Web Page, 20 January 2020) <https://mp.weixin.qq.com/s/fZYa9Zivu7yk0JmbaxoicQ>.

162 Liang Chenyu, 'Shanghai Court Adopts New AI Assistant', *Sixth Tone* (online, 25 January 2019) <sixthtone.com/news/1003496>.

163 Nyu Wang and Michael Yuan Tian, ''Intelligent Justice': AI Implementations in China's Legal Systems' in Ariane Hanemaayer (ed), *Artificial Intelligence and Its Discontents: Critiques from the Social Sciences and Humanities* (Palgrave Macmillan, February 2022) 197, 206-209.

164 Zhuhao Wang, 'China's E-Justice Revolution' (2021) 105(1) *Bolch Judicial Institute* 37, 42 <https://judicature.duke.edu/wp-content/uploads/2021/04/EJustice_Spring2021-1.pdf>.

In February 2023, the Indian Supreme Court implemented AI software for live transcription of court proceedings, marking the first instance of AI being employed in the country's judiciary system.[165] A number of state high courts have started exploring the integration of AI for language translation. For instance, the Kerala High Court is experimenting with an AI-based tool to translate its judgments into Malayalam,[166] while the Delhi High Court has used the SUVAS to publish Hindi versions of some judgments.[167]

Across the border, the Appellate Division of the Supreme Court of Bangladesh introduced a translation software called 'Amar Vasha' in February 2019, which utilises AI to translate court orders and judgments from English to Bangla.[168]

In the Philippines, the Chief Justice Alexander Gesmundo confirmed in April 2023 that the judiciary will integrate AI-enabled voice-to-text transcription services to make stenographic work more efficient and more effective. It forms a part of the Strategic Plan for Judicial Innovations 2022-2027, a blueprint for judicial reform aimed at enhancing the productivity of the courts and freeing up court dockets. The tools will generate automated transcripts, which stenographers can then review, correct, and enhance, and these will be fed back into the software's adaptive algorithms, allowing it to continuously improve.[169]

European countries continue to adopt innovative AI technologies to enhance their judiciaries. Italian courts use technology to create real-time transcripts, but with accountability for accuracy lying with users.[170] State courts in Hungary use automatic speech recognition and transcription systems, which include abbreviations and references to legal rules.[171] Estonia is developing an AI-enabled service that will transcribe court hearings and anonymise judgments.[172] Currently, the Estonian government is testing an online writing tool, called Salme, to expedite the transcription process during court sessions that can accurately document conversations in the courtroom and attributing the statements to their respective speakers with 92% accuracy rate.[173] A text anonymisation initiative – called the MAPA Project – is seeking to introduce a deployable, open-source, multilingual anonymisation toolkit based on natural language processing tools. It is capable of detecting and anonymising personal data in different European languages to help support compliance with data protection regulations.[174] It focuses on de-identification of information from texts, which can be used to anonymise training datasets, court rulings, and other matters required by law.[175]

Chatbot type tools can also be used to provide information to litigants, for example about legal rules and court and registry procedures. The use of generative AI in courtrooms is an emerging area where, to date, there has been more recorded failures than successes. The New York example in section 4.8 is an example of flawed reliance on ChatGPT.

165 Namita Singh, 'Can live-streaming and AI transcriptions restore faith in India's troubled court system?', Independent (online, 24 February 2023) <https://www.independent.co.uk/asia/india/supreme-court-india-judiciary-justice-b2287230.html>; 'Artificial Intelligence is transcribing Supreme Court proceedings: How is it happening, and why?', The Indian Express (online, 22 February 2023) <https://indianexpress.com/article/explained/explained-law/ai-is-transcribing-sc-proceedings-how-is-it-happening-and-why-8458492/>; 'Back Bangalore techies bring AI to Supreme Court for the first time' mint (online, 26 February 2023) <livemint.com/news/india/supreme-court-uses-ai-based-transcript-for-the-first-time-here-s-how-it-works-11677403522929.html>.

166 Mahir Haneef, 'Judgments in Malayalam: Kerala high court tries out AI tool', *The Times of India* (online, 21 February 2023) <timesofindia.indiatimes.com/city/kochi/judgments-in-malayalam-kerala-high-court-tries-out-ai-tool/articleshow/98130202.cms>.

167 'Delhi HC publishes translations of judgments in Hindi', The Times of India (online, 2 March 2023) <https://timesofindia.indiatimes.com/city/delhi/delhi-hc-publishes-translations-of-judgments-in-hindi/articleshow/98348188.cms?from=mdr>.

168 'Supreme Court launches translation software 'Amar Vasha'', *The Business Standard* (online, 18 February 2021) <https://www.tbsnews.net/bangladesh/court/supreme-court-launches-translation-software-amar-vasha-204100>.

169 'SC welcomes use of AI-powered tool', *Manilla Bulletin* (online, 2 May 2023) <https://mb.com.ph/2023/5/1/sc-welcomes-use-of-ai-powered-tools>.

170 Francesco Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings' (2020) 2(1) *Law, Technology and Humans* 4, 12.

171 'Speech Recognition and Transcription Software', *Courts of Hungary* (Web Page) <https://birosag.hu/en/speech-recognition-and-transcription-software>.

172 'Estonia does not develop AI Judge', *EKEI* (Web Page, 11 March 2023) <https://www.ekei.ee/en/news/estonia-does-not-develop-ai-judge>.

173 Martin Hochel, 'Estonian Justice to Be Digitalized With Salme', 3 Seas Europe (Blog Post, 29 March 2023) <https://3seaseurope.com/estonia-justice-salme/>; 'Introducing Salme, Estonian courts' speech recognition assistant', e-Estonia (Web Page, 26 January 2022) <https://e-estonia.com/introducing-salme-estonian-courts-speech-recognition-assistant/>.

174 'MAPA Anonymisation' (Web Page) <https://mapa-project.eu>.

175 Rocío Txabarriaga, 'Unpacking Multilingual Data Management and Anonymization', *slator* (online, 28 September 2022) <https://slator.com/unpacking-multilingual-data-management-and-anonymization/>.

AI language processors often fail to consider the needs of marginalised communities, disregarding their unique ways of engaging with technology. To address this issue, in collaboration with the University of Melbourne, Justice Connect is developing a natural language processing AI model capable of diagnosing legal problems within everyday language.[176] In designing this tool, a specialist team has collaborated with diverse groups, including older individuals, people with disabilities, First Nations people, and individuals from culturally and linguistically diverse backgrounds. Their objective was to gather language samples and gain insights into how people from diverse backgrounds utilize syntax, grammar, shorthand, and slang to describe their legal issues.[177] Having trained the AI model on over 11,000 real-life, de-identified data samples, the diagnostic model is reportedly performing with 88% accuracy across 12 legal category areas.[178]

Judges have expressed different views about the potential of generative AI, and some have been willing to use it openly. In January 2023, a Colombian judge utilised ChatGPT to determine whether an autistic child's insurance should cover the full costs of their medical treatment.[179] While the judgment itself did not generate much controversy, the inclusion of the judge's conversations with ChatGPT in the ruling sparked some debate. The judge had posed the following question: "Is an autistic minor exonerated from paying fees for their therapies?" ChatGPT responded: "Yes, this is correct. According to the regulations in Colombia, minors diagnosed with autism are exempt from paying fees for their therapies."[180] There have been a number of other examples of judges using ChatGPT in the judicial context, including in India in relation to a bail application,[181] in Mexico in relation to electoral law,[182] in Pakistan in relation to bail for a juvenile,[183] and in Peru to calculate child support in a family law case.[184] An English Court of Appeal judge also admitted to using ChatGPT to summarise an area of law.[185]

Although the Colombian judge advocated for the use of generative AI to enhance efficiency, he noted that this technology ought to assist rather than replace human judges. Others are more sceptical. Justice Melissa Perry of the Federal Court of Australia shared her experience with a generative AI tool in a speech at the Commonwealth Law Conference. Justice Perry had asked the tool factual questions about herself and noted the inaccuracy of the response.[186] Unsurprisingly, Justice Perry suggested that the outputs of such

176 Karin Derkley, 'Real life questions to shape AI-powered legal intake tool', *Law Institute Victoria* (online, 15 February 2022) <https://liv.asn.au/Web/Law_Institute_Journal_and_News/Web/LIJ/Year/2022/02February/Real%20_life_questions_to_shape_AI-powered_legal_intake_tool.aspx>.

177 Maggie Coggan, 'The robot-led solution helping marginalised communities find legal help', *PRObono Australia* (online, 1 March 2022) <https://probonoaustralia.com.au/news/2022/03/the-robot-led-solution-helping-marginalised-communities-find-legal-help/>.

178 'Bringing AI to the legal help ecosystem with a free licence for NFPs', *justice connect* (Web Page, 7 March 2023) <https://justiceconnect.org.au/fairmatters/bringing-ai-to-the-legal-help-ecosystem-with-a-free/>.

179 'Colombian judge uses ChatGPT in ruling', *Courthouse News Service* (online, 2 February 2023) <https://www.courthousenews.com/colombian-judge-uses-chatgpt-in-ruling/>; Amy Guthrie, 'Colombian Judge Uses AI Tool ChatGPT in Court Ruling', *ALM Law.com International* (online, 8 February 2023) <https://www.law.com/international-edition/2023/02/08/colombian-judge-uses-ai-tool-chatgpt-in-court-ruling/?slreturn=20230819023611#:~:text=Juan%20Manuel%20Padilla%2C%20a%20judge,should%20receive%20free%20health%20services>.

180 Luke Taylor, 'Colombian judge says he used ChatGPT in ruling', *The Guardian* (online, 3 February 2023) <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling>.

181 'In a first, Punjab and Haryana high court uses Chat GPT to decide bail plea', *The Times of India* (online, 28 March 2023) <https://timesofindia.indiatimes.com/india/in-a-first-punjab-and-haryana-high-court-uses-chat-gpt-for-deciding-upon-bail-plea/articleshow/99070238.cms?from=mdr>.

182 Juan David Gutiérrez, 'Judges and Magistrates in Peru and Mexico Have ChatGPT Fever', *Tech Policy Press* (Blog Post, 19 April 2023) <https://techpolicy.press/judges-and-magistrates-in-peru-and-mexico-have-chatgpt-fever/>.

183 Sandeep Sharma, 'Pakistan court uses ChatGPT to grant pre-arrest bail to a 13-year-old boy who attempted rape', *First Post* (online, 11 April 2023) <https://www.firstpost.com/world/pakistan-court-uses-chatgpt-to-grant-pre-arrest-bail-to-a-13-year-old-boy-who-attempted-rape-12439682.html>; 'In first, Pakistani Judge consults ChatGPT in rape case', *Global Village Space* (online, 12 April 2023) <https://www.globalvillagespace.com/in-first-pakistani-judge-consults-chatgpt-in-rape-case/>.

184 Gutiérrez (n 182).

185 Hibaq Farah, 'Court of appeal judge praises 'jolly useful' ChatGPT after asking it for legal summary', *The Guardian* (online, 15 September 2023) <https://www.theguardian.com/technology/2023/sep/15/court-of-appeal-judge-praises-jolly-useful-chatgpt-after-asking-it-for-legal-summary>; Bianca Castro and John Hyde, 'Solicitor condemns judges for staying silent on 'woeful' reforms'', *The Law Society Gazette* (online, 14 September 2023) <https://www.lawgazette.co.uk/news/solicitor-condemns-judges-for-staying-silent-on-woeful-reforms/5117228.article>

186 Justice Melissa Perry, 'The Future of Administrative Decisions' (Speech, Commonwealth Law Conference, 24 February 2023) <https://www.fedcourt.gov.au/digital-law-library/judges-speeches/justice-perry/perry-j-20230324>.

tools "should be approached with a high degree of caution, if not scepticism".[187] Law firms are more open to experimenting with the appropriate use of generative AI in the context of providing legal services.[188] However, the New Zealand Law Society has warned its members against the dangers of ChatGPT, particularly in the context of invented references.[189]

Generative AI raises other risks in addition to unwarranted reliance on outputs. The confidentiality of inputs may also be at risk depending on the terms and conditions under which a generative AI system is used. For example, if the system is able to learn from inputs, then it is possible that private or confidential information will be recycled indirectly, through the model, into another user's output. There are also broader issues about the compliance of such tools with copyright and privacy laws.

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.[190] 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',[191] including natural language searching and flags for 'bad law'.[192] However, most legal research tools use automation and/or machine learning to help researchers find and link to precedents related to a passage in 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.[193] Lexis Argument Analyser, in particular, uses AI to comb through LexisNexis vast case law database and retrieve relevant case recommendations, analyses of cited cases and legislation, and a map of the salient legal issues, based on a passage from any legal document, uploaded documents, or free-text input.[194] Austlii also uses automation in NoteUp function, which finds documents relevant to the document being viewed.

Expert systems are also used in legal research. An automated tool was designed for Victoria Legal Aid to determine eligibility for legal aid, but never went online.[195] 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.[196] 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.

187 Ibid.

188 Michael Pelly, 'Law firms say ChatGPT an 'opportunity, not a threat'', *Australian Financial Review* (online, 9 February 2023) <https://www.afr.com/companies/professional-services/law-firms-say-chatgpt-an-opportunity-not-a-threat-20230208-p5cj2j>.

189 New Zealand Law Society, 'Beware of legal citations from ChatGPT', *Newsroom* (online, 30 March 2023) <https://www.lawsociety.org.nz/news/legal-news/beware-of-legal-citations-from-chatgpt/>.

190 Jo Szczepanska and Emma Blomkamp, *Seeking Legal Help Online Understanding the 'missing majority'* (Report, November 2020) <https://justiceconnect.org.au/wp-content/uploads/2020/11/Justice-Connect-Seeking-Legal-Help-Online-Missing-Majority-Report-FINAL.pdf>.

191 About Us', *ROSS Intelligence* (Web Page) <https://www.rossintelligence.com/about-us>.

192 'Features', *ROSS Intelligence* (Web Page) <https://www.rossintelligence.com/features>.

193 'The Power of Artificial Intelligence in Legal Research', *LexisNexis* (Web Page, 16 May 2023) <https://www.lexisnexis.com/community/insights/legal/b/thought-leadership/posts/the-power-of-artificial-intelligence-in-legal-research>.

194 'Lexis® Argument Analyser', *LexisNexis* (Web Page) <https://www.lexisnexis.com.au/en/products-and-services/Lexis-Argument-Analyser>.

195 Maria Jean Hall, Andrew Stranieri and John Zeleznikow, 'A Strategy for Evaluating Web-Based Discretionary Decision Support Systems' (Conference Paper, Advances in Databases and Information Systems, 6th East European Conference, 2002) 8-11.

196 See n 28.

3.10 Rules as Code – Implications for the Judiciary

RaC (see section 2.5 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.[197] 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*.[198]

Beyond government, AustLII has developed ElectKB which is essentially a machine-consumable version of section 44 of the Australian *Constitution*.[199] The program is designed as a chatbot to assess whether an individual is eligible to stand as a member of Australia's Federal Parliament.[200]

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.[201] 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.[202]

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.[203] 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.[204]

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 be implemented in the context of urban planning to help inform the new district plan.[205]

197 See 'Community gaming check', *NSW Government Fair Trading* (Web Page) <https://www.fairtrading.nsw.gov.au/community-gaming/community-gaming-regulation-check/>.

198 See 'NSW ESC Estimator', *Nabers* (Web Page) <https://www.nabers.gov.au/rating-tools/our-calculators/nsw-esc-estimator>.

199 James Mohun and Alex Roberts, *Cracking the Code: Rulemaking for Humans and Machines* (OECD Working Papers on Public Governance No 42, Organisation for Economic Co-operation and Development, 2020) 12 <https://www.oecd-ilibrary.org/docserver/3afe6ba5-en.pdf?expires=1695021982&id=id&accname=ocid177499&checksum=5F5F51099CE24BF04D0063A95554EC24>.

200 The chatbot is available at *Constitution Act, s44 consultation* (Web Page) <http://datalex.org/app/consultation?rulebase=http%3A%2F%2Faustlii.community%2Ffoswiki%2FDataLex%2FElectKB>.

201 Mohun and Roberts (n 199) 40-41.

202 Ibid; the English language version of the platform can be accessed at 'The most widely adopted free and open-source engine to write rules as code', *openfisca* (Web Page) <https://openfisca.org/en/>.

203 'Mes Aides', *openfisca* (Web Page) <https://fr.openfisca.org/showcase/mesaides/>.

204 Mohun and Roberts (n 199) 45. One of the simulators is available at *Ma Boussole Aidants* (Web Page) <https://www.maboussoleaidants.fr/mes-aides-financieres>.

205 Mohun and Roberts (n 199) 60; Hamish Fraser, 'What Is Better Rules?', *DIGITAL.GOVT.NZ* (Blog Post, 7 December 2021) <https://www.digital.govt.nz/blog/what-is-better-rules/>.

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.[206] 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.[207] The advantage would be a reduction in compliance risks for organisations seeking to follow the rule, as well as (potentially) higher rates of compliance.[208] 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.[209]

206 Mohun and Roberts (n 199) 44-47.

207 Ibid 80.

208 See 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) 241-243, 247-249.

209 Mohun and Roberts (n 199) 27-28.

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 and independence, 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 tribunals.

4.1 Open Justice

Open justice subjects court proceedings to public and professional scrutiny and is critical to public confidence in the judicial system.[210] 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.[211]

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".[212] 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".[213] 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'"[214] Intentional secrecy produces significant harm to open justice (and accountability, as we discuss below),[215] and "undermine[s] trust in AI and algorithmic outputs".[216] Open justice might also suffer where certain matters (e.g. those conducted through an online platform) are not conducted 'in public'.[217]

210 *Hogan v Hinch* [2011] HCA 4; 243 CLR 506 [20].

211 Susskind was describing first wave AI systems. See Richard E Susskind, *Expert Systems in Law: A Jurisprudential Inquiry* (Clarendon Press 1987) 114–115; Susskind (n 51) 288.

212 *Loomis* (n 92) 774.

213 Ibid 133, 141.

214 Ibid 142.

215 See generally 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. See also Michèle Finck, 'Automated Decision-Making and Administrative Law' in Peter Cane et al (eds), Oxford Handbook of Comparative Administrative Law (Oxford University Press, 2020) 9.

216 Rowland (n 100) 611.

217 Kalliopi Terzidou, 'The Use of Artificial Intelligence in the Judiciary and its Compliance with the Right to a Fair Trial' (2022) 31(3) *Journal of Judicial Administration* 154, 163.

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.[218] 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.[219]

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,[220] and constitute a denial of natural justice.[221]

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 2.12 Explainable AI). 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 2.11 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.

Similar challenges arise when reasons are given for administrative decisions that are automated, in particular where they rely on more complex machine learning algorithms. In 2004, the Administrative Review Council's report on Automated Assistance in Administrative Decision Making concluded that neural networks (an example of machine learning) were not generally suitable for administrative decision-making due to the fact that they "do not easily provide reasons for their decisions".[222] The black box challenge in generating reasons for administrative decisions was addressed more recently by the Commonwealth Ombudsman.[223] They recognised (1) the importance of systems being understandable, so that agencies can demonstrate that decisions are legal and fair, and (2) the importance of statements of reasons, noting that in some cases the output of a system could be edited by human decision-makers to produce a more comprehensible explanation.

Generative AI highlights these various challenges to openness and transparency. While text outputs might appear in the form of reasoning, this may or may not be an actual explanation of 'why' a particular statement was made. Consider the New York example in section 4.8 – ChatGPT stated a legal proposition and suggested a source for that proposition (which might appear as an 'explanation') but the source itself did not exist.

218 See, eg, Wainohu v New South Wales (2011) 243 CLR 181; Henry J. Friendly, '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 at [30] that there is no general common law duty to give reasons in the UK. However, at [41], the Court accepted that a duty to give reasons could arise in some circumstances, presumably alluding to situations in which procedural fairness would require it.

219 Finck (n 215) 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, 870; Rowland (n 100) 602 ('Its logic can be made more transparent and easier to adjust than human thinking').

220 Richard M Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22 *Stanford Technology Law Review* 242, 264–5.

221 Danielle Keats Citron, 'Technological Due Process' (2008) 85(6) Washington University Law Review 1249, 1252.

222 Administrative Review Council, *Automated Assistance in Administrative Decision Making: Report to the Attorney-General* (Report No 46, November 2004), section 2.5 <https://www.ag.gov.au/sites/default/files/2020-03/report-46.pdf> ('*Administrative Law Council Report*').

223 Commonwealth Ombudsman, *Automated Decision-making: Better Practice Guide* (February 2007, as updated in 2019), 25-26 <https://www.ombudsman.gov.au/__data/assets/pdf_file/0029/288236/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf> ('*Commonwealth Ombudsman Guide*').

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 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.[224] 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.[225]

THINGS TO CONSIDER – Questions for courts and tribunals

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:
    - **a** lack of disclosure by the provider of the AI system;
    - **b** contractual promises to keep information confidential;
    - **c** privacy of personal information in training data;
    - **d** information being provided in a form that cannot be understood by the relevant audience (for example, as computer code);
    - **e** information being provided about the functioning of the system that is too complex for a functional understanding of the reasons for a given output;
    - **f** information being provided about the system that may not be a true representation of the operation of the system or 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)?

224 Re and Solow-Niederman (n 220) 262–267; Andrew D Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) 87 *Fordham Law Review* 1085, 1099.

225 Leah Wisser, 'Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing' (2019) 56(4) *American Criminal Law Review* 1811, 1815.

4.2 Judicial Accountability and Independence

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.[226] 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. Similarly, judicial independence – the capacity of the courts to perform their function free from interference and dependence upon any persons or institutions, including the executive arm of government, over which they do not exercise direct control[227] – may be impinged where the AI tool is chosen, operated or owned by a third party.

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. The way in which any 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.[228] 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.[229] Where AI systems are used in the courtroom or tribunal hearing, 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.

226 Selbst and Barocas (n 224) 1088.

227 Encyclopaedic Australian Legal Dictionary (LexisAdvance, Online) 'judicial independence'; Guy Green, 'The Rationale and Some Aspects of Judicial Independence' (1985) 59 *Australian Law Journal* 135, 135.

228 Bloch-Wehba (n 232) 1308; *An Act relating to establishing guidelines for government procurement and use of automated decision systems in order to protect consumers, improve transparency, and create more market predictability*, HB 1655 - 2019-20, Wash 66th Legislature (2019).

229 Finale Doshi-Velez et al, 'Accountability of AI Under the Law: The Role of Explanation' (Working Paper, Berkman Klein Center Working Group on Explanation and the Law, 2017) 16–17.

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".[230] 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.*[231]

Similar cautions around basing decisions solely on automated processes can be found in Australian and New Zealand policy. New Zealand has an Algorithm Charter, being a commitment by government agencies to manage the use of 'algorithms' in a fair, ethical and transparent way.[232] The Australian Administrative Law Council in its 2004 report included principles that expert systems (the most relevant technology at the time) should not be used to automate the exercise of discretion but could be used to assist a human officer in the exercise of discretion.[233] The Commonwealth Ombudsman in its report also recognised the importance of transparency and accountability in the use of automated decision-making, including specific recommendations around transparency and publicly available information about systems, understandability of systems, audit, reasons for decision, review of decisions, and monitoring and evaluation.[234] New South Wales has an AI Assurance Framework for the public service when it deploys, uses or manages AI systems, including specific risk-based questions around transparency and accountability.[235] The Commonwealth Department of Industry, Science and Resources has also published an ethical framework for AI, a framework revolving around high level principles including accountability and transparency.[236] Unlike the New South Wales framework, this is not limited to government and is entirely voluntary, as well as providing significantly less detail.

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".[237] If the courts increasingly implement AI tools which judges may not themselves understand, how can they remain accountable to the public? For example, the proposed EU Artificial Intelligence Act ('AI Act') classifies AI tools used in the administration of justice as high-risk.[238] The Article 8 of the Annex III of the AI Act defines AI systems broadly: tools intended to assist judicial authorities to conduct research, interpret and apply the law to a set of circumstances.[239] All such AI tools will be subject to legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security.[240]

230 *GDPR* (n 31) art 22(1).

231 Ibid art 4(4).

232 See The Algorithm Charter for Aotearoa New Zealand (July 2020) <https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/>. See also The Law Foundation New Zealand and University of Otago, Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project (2019) <https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf>.

233 *Administrative Law Council Report* (n 222), Principles 1, 2, 3.

234 *Commonwealth Ombudsman Guide* (n 223), 25-27.

235 See NSW Government, *Artificial Intelligence assurance framework* (September 2022) <https://www.digital.nsw.gov.au/sites/default/files/2022-09/nsw-government-assurance-framework.pdf>.

236 See Australian Government Department of Industry, Science and Resources, *Australia's Artificial Intelligence Ethics Framework* (7 November 2019) <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework>.

237 House of Lords, *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].* (Report, March 2017) <https://biginnovationcentre.com/wp-content/uploads/2023/05/BIC_APPG-AI-THEME-REPORT-EVIDENCE-MEETING-2-ETHICS-AND-LEGAL-IN-AI-DECISION-MAKING-AND-MORAL-ISSUES_20.09.2017-1.pdf>.

238 Recital 40 of the Proposed AI Act reads: "Certain AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. In particular, to address the risks of potential biases, errors and opacity, it is appropriate to qualify as high-risk AI systems intended to assist judicial authorities in researching and interpreting facts and the law and in applying the law to a concrete set of facts. Such qualification should not extend, however, to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks or allocation of resources."

239 See *Proposed AI Act* (n 8) annex III, art 8(a).

240 Ibid Chapter 2, Title III.

In the proposed EU AI Act, the role of the AI system providers and controlling/compliance/testing is particularly important for judicial values, in particular judicial independence and accountability. For example, Article 3 of the AI Act defines a provider as "a natural or legal person, public authority, agency or other body" that develops an AI system[241] and this may include the executive and legislative bodies. Judicial accountability and other values would be undermined if national governments designed algorithms used in courts. Similarly, influence from public authorities to the judiciary can be indirect via private companies developing AI tools. Also, there are risks of potential control, interference and surveillance from foreign states via privately developed AI tools.[242] Recent Pegasus[243] and PRISM[244] scandals illustrate the reality and frequency of this risk. This is particularly significant for judicial accountability for AI-assisted courts, which would undermine public confidence and trust in judicial systems.[245]

The proposed AI Act will also place AI tools used by the judiciary under control of many entities and this could impact on judicial independence as well as judicial accountability. In particular, the AI Act lists numerous bodies to be in charge or responsible for, *inter alia*, conformity assessment of AI systems,[246] carrying out necessary procedures for the assessment, designation and notification of conformity assessment.[247] It also elaborates on the role of the European Artificial Intelligence Board, and national competent authorities.[248] As Giulia Gentile notes, while the EU *AI Act* stipulates some requirements of independence for some of these bodies, they do not seem to cover national competent authorities.[249] Thus, it is unclear how the interaction of various bodies charged with overseeing the AI tools in the courtrooms with the executive and the legislator impact on judicial independence and accountability.

The EU has also adopted the e-Justice Strategy and Action Plan 2019-2023[250] which has identified the use of AI and blockchain/DLT in the justice field as priority. The *EU Study* on the Use of Innovative Technologies was released in 2020, demonstrating that there are numerous projects in the EU Member States with similar objectives, and technologies, and arguing for better coordination of efforts and activities at EU-level.[251]

In Australia, AI systems are governed by general law, rather than technology-specific laws, although there are proposals to focus more directly on regulation of AI as such. The advantages and disadvantages of such an approach will not be canvassed here,[252] but it is useful to outline where things currently stand. There are also a series of policy reports addressing the question of legislation, regulation and ethical practice for AI, whether as technology-specific law, broader law reform, or proposed guidance, that could apply to public and private sectors.[253] Even before any new laws or reforms, however, there are a number of laws that will be applicable to AI systems *inter alia*. Courts and tribunals, as arbiters of the law, are accountable for the legality of their processes. Some AI systems may fall foul of Australian privacy law or Australian copyright law, and

241 Ibid art 3.

242 Ibid recital 10.

243 Stephanie Kirchgaessner et al, 'Revealed: leak uncovers global abuse of cyber-surveillance weapon', *The Guardian* (online, 19 July 2021) <https://www.theguardian.com/world/2021/jul/18/revealed-leak-uncovers-global-abuse-of-cyber-surveillance-weapon-nso-group-pegasus>.

244 Glenn Greenwald and Ewen MacAskill, 'NSA Prism program taps in to user data of Apple, Google and others', *The Guardian* (online, 8 June 2013) <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>.

245 Giulia Gentile, 'AI in the courtroom and judicial independence: An EU perspective', *EUIdeas* (online, 22 August 2022) <https://euideas.eui.eu/2022/08/22/ai-in-the-courtroom-and-judicial-independence-an-eu-perspective//>.

246 *Proposed AI Act* (n 8) art 33.

247 Ibid art 30.

248 Ibid art 59.

249 Gentile (n 245).

250 *2019–2023 Action Plan European e-Justice* [2019] OJ C 96/05/9 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019XG0313(02)&rid=6>.

251 Miglena Vucheva et al, *The Study on Innovative Technologies in the Justice Field* (Final Report, September 2020) <https://op.europa.eu/en/publication-detail/-/publication/4fb8e194-f634-11ea-991b-01aa75ed71a1/language-en>.

252 See Lyria Bennett Moses 'Regulating in the Face of Socio-Technical Change', in Roger Brownsword, Karen Yeung and Eloise Scotford (eds.), *Oxford Handbook of the Law and Regulation of Technology* (Oxford: Oxford University Press, 2017).

253 See generally Australian Law Reform Commission, *The Future of Law Reform: A Suggested Program of Work 2020–25* (Report, 2 December 2019) <https://www.alrc.gov.au/wp-content/uploads/2019/11/Future-of-Law-Reform-Final-Report_v3web.pdf>; Genevieve Bell et al (n 15) (n; *Safe and responsible AI Discussion Paper* (n 7); Parliament of New South Wales, Inquiry on Artificial Intelligence (AI) in New South Wales (Web Page, 27 June 2023) <https://www.parliament.nsw.gov.au/committees/inquiries/Pages/inquiry-details.aspx?pk=2968>; Parliament of South Australia Select Committee on Artificial Intelligence (Web Page, 11 July 2023) <https://www.parliament.sa.gov.au/en/News/2023/07/11/03/37/Select-Committee-on-Artificial-Intelligence>.

this should be checked prior to their use. By way of example, the Australian Federal Police were criticised by the Office of the Australian Information Commissioner for its use of Clearview AI, which relied on data scraping in building a facial recognition service.[254]

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.[255] 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 AI and automation in public decision-making more broadly.

THINGS TO CONSIDER – Questions for courts and tribunals

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?
5. Are there reasons why it may be illegal or inappropriate to use the system, either generally or in the context proposed? What might be the consequences for real and perceived independence and accountability?

254 See Office of the Australian Information Commissioner, *Commissioner initiated investigation into Clearview AI, Inc. (Privacy)* [2021] AICmr 54 (14 October 2021).

255 Bennett-Moses, Boughey and Crawford (n 208) 240; 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.[256] 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.

Bias occurs not only in classification systems (such as COMPAS which classifies people based on risk level) but also in generative AI systems. For example, the system may make assumptions as to the gender of a person described in the input neutrally as "judge" based on an association drawn from the material on which it was trained. AI 'detector' tools (which seek to predict whether text was authored by AI) are more likely to flag text written by those writing in a non-native language.[257] More broadly, AI systems trained on data that comes primarily from US jurisdictions may have a bias towards answering questions based on patterns identified in US texts, using terms and concepts which may not be applicable to Australia.

Humans are also prone to bias. Sometimes this is overt but, more often, it is unconscious.[258] 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.[259] 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.[260]

256 Ozkan Eren and Naci Mocan, 'Emotional Judges and Unlucky Juveniles' (2018) 10(3) *American Economics Journal: Applied Economics* 171, 174-175, 200-203.

257 Weixin Liang et al, 'GPT detectors are biased against non-native English writers' (2023) 4(7) *Patterns* 100779.

258 You can take a test to explore your implicit biases at *Project Implicit* (Web Page) <https://implicit.harvard.edu/implicit/takeatest.html>.

259 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 18) 96-97.

260 Felicity Bell, 'Family Law, Access to Justice, and Automation' (2019) 19 *Macquarie Law Journal* 103, 118.

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.[261] 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.[262] 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".[263] Replacing or supplementing discretion with AI systems is delegating "some of our moral responsibility",[264] 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 risk scores.[265] 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.[266] At the very least, we could expect risk scores to generate a framing or anchoring effect.[267]

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.[268]

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.[269] 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.

261 Ward (n 156).

262 Rebecca Crootof, '"Cyborg Justice" and the Risk of Technological–Legal Lock-In' (2019) 119(1) *Columbia Law Review* 1, 4. See also 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.

263 Stowe Boyd, 'A Speculative Post on the Idea of Algorithmic Authority « Clay Shirky', *Medium* (online, 23 November 2009) <https://stoweboyd.medium.com/a-speculative-post-on-the-idea-of-algorithmic-authority-clay-shirky-c248019a0921>.

264 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.

265 Carlson (n 91) 319-320.

266 *Injustice Ex Machina* (n 100), citing 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.

267 Todd McElroy and David Dickinson, 'Thoughtful Days and Valenced Nights: How Much Will You Think about the Problem?' (2010) 5(7) *Judgment and Decision Making* 516, 516.

268 Re and Solow-Niederman (n 220) 264, 266–7.

269 *Wong v The Queen* (n 113) [65]; *Green v The Queen* (2011) 244 CLR 462, 472–3 [28].

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.[270] As described in section 3.3 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.[271] 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'"[272]

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,[273] "[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".[274]

Australia is currently considering how AI and other technological tools can impact on judicial impartiality in ongoing federal reform;[275] 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.[276]

270 Pasquale and Cashwell (n 75) 80-81.

271 See *BDS17 v Minister for Immigration and Border Protection* [2018] FCA 1683.

272 *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-analytics5-years-in-prison-for-rule-breakers/>.

273 Bathurst (n 116) [54].

274 Ibid.

275 'Review of Judicial Impartiality', Australian Law Reform Commission (Web Page, 11 September 2020) <https://www.alrc.gov.au/inquiry/review-of-judicial-impartiality/>.

276 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.

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.

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".[277] 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:[278]

> *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.*

277 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); *HT v The Queen* [2019] HCA 40, [17] (Kiefel CJ, Bell and Keane JJ), [64] (Gordon J).

278 *International Finance Trust Co Ltd v New South Wales Crime Commission* (2009) 240 CLR 319, 379 [141].

Procedural fairness may also include the parties being given an opportunity to call their own witnesses and to cross-examine the opposing witnesses.[279] An opponent may not advance contentions or adduce evidence of which a party is kept in ignorance.[280] 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?

Litigants, particularly in criminal contexts, should be placed in a position to challenge evidence against them. Classifications, predictions and AI-generated content might be introduced into evidence, either with or without the judge (or jury) being aware of the involvement of AI. An image that appears to be a photograph of an accused doing something could be a 'deep fake' created using AI. An AI system may be able to generate text that mimics the way in which a litigant usually expresses themselves. Demonstrating, for example, that some piece of evidence is a fake or a forgery may become increasingly difficult as the technology develops. Litigants may thus require greater resources to challenge evidence.

The outputs of risk assessment tools may be used directly (as where the output is introduced into evidence in sentencing) or indirectly (as where an expert relies on the tool in producing his or her report). 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] [2018] WASC 310, 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.[281] Historic versions of systems should also be archived.

279 *Al Rawi v Security Service* [2011] UKSC 34; [2012] 1 AC 531 [13].

280 Ibid [12].

281 Ombudsman New South Wales, The New Machinery of Government: Using Machine Technology in Administrative Decision-Making (Report, 29 November 2021) 48 <https://www.ombo.nsw.gov.au/__data/assets/pdf_file/0003/138207/The-new-machinery-of-government-special-report_Front-section.pdf>.

THINGS TO CONSIDER – Questions for courts and tribunals

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 to 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.[282] Self-represented litigants may be particularly interested in using AI tools and ODR systems to assist them in preparing documents and understanding and managing the steps involved in their dispute.[283] 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.[284] 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.[285]

282 See, eg, 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).

283 See generally Amy J Schmitz and John Zeleznikow, 'Intelligent Legal Tech to Empower Self-Represented Litigants' (2021) XXIII *Columbia Science & Technology Law Review* 142.

284 Legg and Bell (n 18) 140-143.

285 Victor Li and Sean La Roque-Doherty, 'Towards Smarter Courts: Artificial Intelligence Has Made Great Inroads, but Hasn't yet Increased Access to Civil Justice', (2021) 107(2) *ABA Journal* 20, 20.

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.[286] After an announcement in early 2023 that the chatbot will provide real-time instructions to a defendant in judicial proceedings, objections to the AI-assisted defence and threat of legal action led to the discontinuation of the experiment.[287]

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 Litigation Outcomes) could be used to triage matters and prioritise those most likely to succeed.[288] This could impinge on a person's right to be heard (see previous section 4.4 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.[289] Similarly, human experience and discretion is central to most judicial decisions and cannot be meaningfully exercised by any known AI systems.[290] More broadly, losing judges' emotion, morality, indeterminacy and creativity would fundamentally change what justice looks like.[291] 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 and tribunals

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?

286 'Save Time and Money with DoNotPay!' <https://join.donotpay.com>.

287 @jbrowder1 (Joshua Browder) (X, 21 January 2023) <https://twitter.com/jbrowder1/status/1616628244840579074> stating: "On February 22nd at 1.30PM, history will be made. For the first time ever, a robot will represent someone in a US courtroom. DoNotPay A.I will whisper in someone's ear exactly what to say. We will release the results and share more after it happens. Wish us luck!"; Megan Serullo, 'AI-powered "robot" lawyer won't argue in court after jail threats', *CBS News* (online, 26 January 2023) <https://www.cbsnews.com/news/robot-lawyer-wont-argue-court-jail-threats-do-not-pay/>.

288 Aletras et al (n 73).

289 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.

290 Meena Hanna, 'Robo-Judge: Common Law Theory and the Artificially Intelligent Judiciary' (2019) 29 *Journal of Judicial Administration* 22, 39–42.

291 Bathurst (n 116).

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.[292] The link between efficiency and AI is clear; as judges themselves are both an expensive and limited resource,[293] automated decision-making has been suggested as a cheap, fast and scalable alternative.[294] However, efficiency comes with 'implied strings'[295] or 'trade-offs',[296] 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[297] and the use of e-filing systems (see section 3.6 Automated E-Filing)[298] 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.[299]

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.

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.

292 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, 169-170.

293 Crootof (n 263), noting that 'human judges are an inherently expensive and limited resource: They must prepare for years, they take time to decide cases, and they retire'.

294 Re and Solow-Niederman (n 220) 287.

295 Jennifer Walker Elrod, 'Trial by Siri: AI Come to the Courtroom' (2019) 57 *Houston Law Review* 1083, 1093.

296 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.

297 Sheryl Jackson, 'Court-Provided Trial Technology: Efficiency and Fairness for Criminal Trials' (2010) 39(3) *Common Law World Review* 219, 236-7.

298 HMCTS E-Filing service for citizens and professionals', GOV.UK (Web Page, 30 September 2022) <https://www.gov.uk/guidance/hmcts-e-filing-service-for-citizens-and-professionals>;'Electronic Filing (CM/ECF)', United States Courts (Web Page, October 2021) <https://www.uscourts.gov/court-records/electronic-filing-cmecf>; Australian National Court Framework', *Federal Court of Australia* (Web Page12 November 2019) <https://www.fedcourt.gov.au/about/national-courtframework> .

299 Kim Neeson, 'Is AI Coming to a Legal Transcript Near You?', *LexisNexis Canada* (Web Page, June 2019) <https://www.lexisnexis.ca/en-ca/sl/2019-06/is-AI-coming-to-a+legal-transcript-near-you.page>.

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. In the UK, the use of pre-approved divorce software generated an error, requiring 2,235 cases to be reopened and information to be resubmitted.[300] Given the cost involved in creating AI tools to begin with, there ought to be sharply superior results for the attendant difficulties (e.g. opacity, possible bias) to be worthwhile.

Finally, Reichman et al, in an analysis of Israel's Legal-Net system,[301] 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".[302] 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 and tribunals

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 RaC 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 RaC 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.[303] Nevertheless, the closer the proposed use is to the core of judicial decision-making, the more caution is required.[304]

300 Francesco Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings' (2020) 2(1) Law, Technology and Humans 4, 9-10.

301 Reichman, Sagy and Balaban (n 143).

302 Ibid 635.

303 Sourdin (n 11) 235.

304 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.

4.8 Court and Tribunal Responses to the Use of AI by Lawyers

There has been some cases where the use of AI in *courts and tribunals* is litigated directly. This is in addition to cases where AI is part of the subject matter or context for a dispute as with questions of whether AI can be an 'inventor' for a patent, whether generative AI models infringe copyright, liability for harm caused by AI systems, and so forth.

One example from New York was disciplinary proceedings concerning inappropriate reliance on ChatGPT.[305] The lawyers concerned used ChatGPT as a legal research and brief drafting tool, apparently unaware of its tendency to 'hallucinate' and make things up, including case law. They were representing a plaintiff, who alleged that he was injured when a metal serving cart struck his left knee during a flight from El Salvador to the John F. Kennedy Airport. The defendant sought to strike the claim out on the basis that the claim was statute-barred. The plaintiff's lawyers asked ChatGPT to provide case law that would support opposing the dismissal. ChatGPT produced a number of cases which the lawyers included in their brief to the Court without reviewing the actual cases. None of the cases existed. When asked by the lawyer whether the cases were genuine, ChatGPT replied in the affirmative, and was even able to produce (fake) extracts from the cases.[306] This makes sense in the context of predictive text – the fake cases and references were similar in format to genuine cases. But reliance on the outputs of generative AI as doctrinally meaningful statements ignores the limitations of generative AI tools such as ChatGPT.

The lawyers and their law firm were subject to court-imposed sanctions of $5,000, pursuant to the Federal Rules of Civil Procedure r 11 or the court's inherent authority, for submitting a brief with fake cases made up by ChatGPT and then standing by the research.[307] United States District Judge Castel also required that the lawyers write to each judge falsely identified as the author of the fake cases attaching his Honour's judgment and orders. His Honour stated that there "is potential harm to the reputation of judges and courts whose names are falsely invoked as authors of the bogus opinions and to the reputation of a party attributed with fictional conduct". [308]

Some jurisdictions have issued guidelines about the use of artificial intelligence in general and generative artificial intelligence in particular. Shortly before this guide went to print, we noted the United Kingdom Courts and Tribunals Judiciary, Artificial Intelligence (AI) Guidance for Judicial Office Holders (12 December 2023) and the Courts of New Zealand's Guidelines for use of generative artificial intelligence in Courts and Tribunals (7 December 2023), with separate guidelines for (1) judges, judicial officers, tribunal members and judicial support staff, (2) lawyers, and (3) non-lawyers (including self-represented litigants).

305 Kathryn Armstrong, 'ChatGPT: US lawyer admits using AI for case research', *BBC* (online, 27 May 2023) <https://www.bbc.com/news/world-us-canada-65735769>; Larry Neumeister, 'Lawyers blame ChatGPT for tricking them into citing bogus case law', *AP* (online, 9 June 2023) <https://apnews.com/article/artificial-intelligence-chatgpt-courts-e15023d7e6fdf4f099aa122437dbb59b>.

306 Benjamin Weiser, 'Here's What Happens When Your Lawyer Uses ChatGPT', The New York Times (online, 27 May 2023) <https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html>

307 *Mata v Avianca Inc* (SDNY, 22-cv-1461 (PKC), 22 June2023) ('*Mata v Avianca*'). See also Debra Cassens Weiss, 'Lawyers who 'doubled down' and defended ChatGPT's fake cases must pay $5K, judge says' *ABA Journal* (online, 26 June 2023) <https://www.abajournal.com/web/article/lawyers-who-doubled-down-and-defended-chatgpts-fake-cases-must-pay-5k-judge-says#:~:text=A%20federal%20judge%20in%20New%20York%20City%20has%20ordered%20two,then%20standing%20by%20the%20research>.

308 *Mata v Avianca* (n 308) 2, 34.

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 or tribunals, members of the judiciary or tribunal, 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. That task will not always be easy as most of the judicial values discussed cannot be converted into the kinds of precise, mathematical requirements that are easy to embed directly into technical systems.[309] Ultimately, each AI system is different – so it is necessary to ask *each time* whether, in relation to a *particular* system, there are *particular* concerns which could jeopardise the open, accountable, impartial, fair and efficient delivery of justice. Kroll has suggested that technical experts be appointed as 'special masters' in US courts to evaluate software-driven decision-making tools.[310] He also discusses advantages and limitations of various testing methodologies.[311] 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. Continuous monitoring and regular evaluation will also assist in checking that behaviour corresponds to expectations in this regard.

An alternative approach is to develop a set of rules or principles for the use of AI systems in courts. For example, in Brazil, in 2020, the National Council of Justice (CNJ) published *Resolution n 332/2020 on the use of (AI) by the Judiciary*.[312] The Resolution establishes ethical governance criteria to guide the development of AI systems, so that they are aligned with the protection and guarantee of fundamental rights, data protection and privacy. Similarly, in Europe, the *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*,[313] sets out principles, such as compatibility with fundamental rights, that systems must satisfy. The document, prepared by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe in 2018, is non-binding and elaborates five core principles to guide both policy makers, legislators and legal professionals, among other stakeholders, working with judicial AI tools. In the European Union, several prominent legislative proposals have been tabled in recent years that will have direct impact on the use of AI by the judiciary. For example, the latest proposed AI Act, and the Data Governance Act[314] accompany EU declaratory instruments, such as its recent declaration of rights and principles in the digital age, adopted by the EU Commission in 2022.[315] There are also calls in the US for

309 Christoph Winter, Nicholas Hollman and David Manheim, 'Value Alignment for Advanced Artificial Judicial Intelligence' (2023) 60(2) American Philosophical Quarterly 187, 191.

310 Joshua A. Kroll et al, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633, 655.

311 Ibid.

312 National Council of Justice, Resolução No 332, De 21 De Agosto De 2020 [Resolution No. 332, 20 August 2020] (Brazil) 20 August 2020 <https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf>.

313 European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment, adopted at the 31st plenary meeting of European Commission for the Efficiency of Justice (Strasbourg, 3-4 December 2018) <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

314 Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act), COM (2020) 767 final (25 November 2020).

315 See Commission Proposal for a Declaration on Digital Rights and Principles for the Digital Decade, COM (2022) 28 final (26 January 2022).

better governance in this area.[316] Although it deals with only one aspect of AI, in May 2023, a judge on the U.S. District Court of the Northern District of Texas implemented a new requirement for attorneys appearing before him to certify that either no part of the document was generated by an AI tool like ChatGPT, or that a human has reviewed any AI-generated text. Any filings lacking the required sworn attestation will be rejected by the court.[317]

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 KC asks whether individuals will feel they are treated fairly in their interaction with the legal system.[318] Will computer outputs warrant or receive the same respect as human judges?[319] 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.[320]

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

- *Why* is AI being used? What problem does it solve?
- Is the use of AI authorised in the context in which it is deployed?
- *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?
- *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*? Will the use of AI in courtrooms and tribunals be accepted by the public?

316 Carlos E. Jimenez-Gomez and Jesus Cano Carrillo, 'Essential Elements and Ethical Principles for Trustworthy Artificial Intelligence Adoption in Courts' in Charles Campbell and John Holtzclaw (eds), *Trends in State Courts 2022* (National Center for State Courts, 2022) 119, 126-128.

317 Jacqueline Thomsen, 'US judge orders lawyers to sign AI pledge, warning chatbots 'make stuff up'', *Reuters* (online, 3 June 2023) <https://www.reuters.com/legal/transactional/us-judge-orders-lawyers-sign-ai-pledge-warning-they-make-stuff-up-2023-05-31/>; Megan Cerullo, 'Texas judge bans filings solely created by AI after ChatGPT made up cases', *CBS News* (online, 2 June 2023) <https://www.cbsnews.com/news/texas-judge-bans-chatgpt-court-filing/>.

318 Margaret Beazley, 'Law in the Age of the Algorithm' (Speech, State of the Profession Address to New South Wales Young Lawyers, 21 September 2017) [64] <https://supremecourt.nsw.gov.au/documents/Publications/Speeches/2017-Speeches/Beazley_20170921.pdf>.

319 See also Sourdin (n 11).

320 Legg (n 293) 181-183.

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);
a. Scale, 1 (not confident) – 5 (very confident)

b. 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)

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

b 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;

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

   b Prediction of judicial outcomes more broadly;

   c Generation of litigation strategies;

   d Natural language processing for translation and transcription purposes;

   e Automated electronic filing;

   f Triaging and allocation of matters;

   g Judicial administration and judicial metrics;

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

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

   j Automated adjudication of matters with consent of the parties;

   k Automated online dispute resolution;

   l Risk assessment tools in sentencing;

   m 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?

   a Open justice and transparency

   b Judicial accountability

   c Judicial independence

d Judicial impartiality

e Procedural fairness

f Access to justice

g Economy and efficiency

h Constitutionality

i 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?

a Ensuring that humans are involved in some stage of the automated processes.

b Ability to appeal to a human decision-maker.

c Preventing the underlying algorithm being covered by trade or state secrecy laws, or other intellectual property protections:

d Ensuring that some form of reasons are given.

e Ensuring that the system developers are accountable for the output of an AI system.

f Ensuring that the AI system produces demonstrably equal outcomes in relation to protected characteristics including age, disability, race, religion and sex.

g Ensuring that there is no risk of tampering with or 'hacking' the AI system.

2 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

3 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

1 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.

a Free text

www.aija.org.au