

Legal Futures: Where Should Artificial Intelligence Take Us?

Professor Lyria Bennett Moses, UNSW*

Abstract

It is never wise to predict the future – evolving capabilities of artificial intelligence systems in prediction and content generation are impressive and sometimes surprising. The original goal of artificial intelligence research was to create machines that simulated human learning and intelligence. In the outputs of generative AI, we can indeed find simulated human reasoning and text that could have been written by a human. Lawyers, alongside many others, have been impressed with AI systems' ability to answer questions, pass exams, and produce documents that mirror, in at least some ways, the performance of human law students and lawyers. In light of this, it is tempting to reduce recruitment and start contemplating a world of avatar judges delivering automated justice based on documents authored through the use of large language models fine-tuned on the wealth of legal documents held in large law firms and the corpus of statutes and case law.

Before rushing towards such a future, we should pause to consider both what might be gained (and the skills we will need to achieve that) as well as what would be lost. In doing this, we need to step back and understand how different kinds of artificial intelligence system work – what they do well and where they can fail. Then, projecting into the future, we can ask about the affordances and limitations of hypothetical developments in artificial intelligence. Only after that can we approach some important questions: What skills will we need to provide legal services and operate legal institutions optimally in light of the capabilities afforded by developments in artificial intelligence? And, more importantly, how do we ensure that the core purposes and functions of our legal system are preserved by recognising the limits of simulation?

Technologies of law

This talk is about the future of law. I am going to look at artificial intelligence and how it is and will change what legal work looks like and how legal institutions operate. I am not only going to talk about data-driven prediction and large language models but Neuralink, uploaded minds and star trek characters. But, before we get there, I need to start with some history. That way, we can use the past to understand the contingency of the present and our choices for the future.

Prior to colonisation, the First Nations of Australia had systems of law that involved a high degree of human mediation and an oral tradition that was closely aligned with lore and culture.¹ Indigenous peoples communicate such traditions through traditional means such as songlines and art, or nowadays through written text, in ways that are appropriate and adapted to their cultures.

*Director, UNSW Allens Hub for Technology, Law and Innovation

¹ Marcia Langton and Aaron Corn, *Law: The Way of the Ancestors* (Thames & Hudson Australia, 2023).

Turning to a different part of the world, this picture shows the Code of Hammurabi, established between 1792 and 1750 BCE.



The laws were inscribed on a stone pillar and displayed in public. The technology of stone engraving was excellent for preservation and works well for a society that believes that law is divinely inspired and unchanging. It reflects and creates a power structure, favouring literate elites living in centres where the law was installed.² It would, however, be impracticable for the current, growing, and rapidly changing statutory corpus of a modern country.

Post-colonial Australian law was for most of its life based on the technology of printed script.³ This combines the affordances of writing with greater diffusion than would be possible with stone tablets. First, moving to paper facilitates rule over larger, more geographically diffuse communities than could be accomplished by reliance on an oral tradition or engraved stone. Geographical spread and large domains of application also separates the creator of text from readers, bringing into play the need for a process of interpretation that does not involve direct questions to the text's author.⁴ Such interpretations become part of the activity around the law, also captured in text. Printing allows such material to diffuse throughout regions subject to the same laws, and even different jurisdictions with similar legal traditions.

More recently, digitalisation has led to a growth in the volume of both legal and factual material available. Doing research in contexts such as decision-making, deal-making and dispute resolution now involves far more materials than might have been selected and published. Both primary and secondary material are also made available more quickly, meaning that up-to-dateness is counted in days rather than months. For example, I used to spend school holidays putting stamps or stickers on reported cases that pointed to newer cases in which those cases were cited. Waiting for the stickers to be printed and for school holidays for the availability of child labour was not seen as a problem.

² Mireille Hildebrandt, 'A Vision of Ambient Law' in Roger Brownsword and Karen Yeung (eds), *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes* (Hart, 2008) 175, 179–180.

³ Hildebrandt (n 2).

⁴ *Ibid* 181.

With digitalisation, we now automatically manage fast paced change in the corpus of legal materials. Digitalisation has also increased our ability to copy, amend and share documents. All of this has increased the importance and complexity of legal information management and systematisation.⁵

These developments are mostly positive, particularly in the context of a complex modern society such as Australia. It is worth noting, however, that there are some disadvantages and limitations of digital technologies. Some have suggested that typing rather than writing changes the parts of the brain involved, which may lead to more subtle differences in legal reasoning as a result of digitalisation.⁶ Mostly, however, the limitations of digitalisation only move to the front of our minds when the Internet is unavailable, when critical computer networks are hacked, or when a system's functionality is impaired. A hidden affordance of law being in physical books, such as the Commonwealth Law Reports, is that the integrity of the reports was easy to maintain. To change the law, a hacker would have to break into every lawyer's office and substitute a fake legislation or report volume. By centralising storage of legal databases, we rely on platforms to ensure the integrity and availability of their files.

The lesson from all of this is that technologies, including legal technologies, might be good or bad in different ways but are never neutral.⁷ Technology not only changes how law is practised, but it also changes the way that law operates in society. Technologies are, like the law itself, normative.⁸ If we want to understand the implications of artificial intelligence in law, we first need to understand how it works.

Artificial intelligence as legal technologies

Technology as applied to law is an interdisciplinary space. Those working in data science and artificial intelligence often assume that their methods transcend domains and can be applied equally to business, law, medicine and other fields. Lawyers are similar, assuming that they can analyse law as applied across diverse industries and factual contexts. In the case of legal technology, both groups risk falling into a trap of assuming that they understand the other's domain. The way law is practiced and the way it operates in society is too important to leave to technologists,⁹ but when lawyers comment on the appropriateness of using particular tools, they first need to understand those tools. Vague handwaving about a 'singularity' or 'general' artificial intelligence is insufficient.¹⁰ Asking whether 'technology' can replace humans is meaningless unless one has in mind both a specific application and a specific vision for the task being done.¹¹

Artificial intelligence has not had a consistent definition over time. Its changing meaning is also rapid, as illustrated by the change in the OECD definition of 'AI system' between 2019 and 2023:¹²

⁵ Ibid 182–183.

⁶ Maria Konnikova, 'What's Lost as Handwriting Fades', *New York Times* (online, 2 June 2014) <<https://www.nytimes.com/2014/06/03/science/whats-lost-as-handwriting-fades.html>>.

⁷ Melvin Kranzberg, 'Technology and History: "Kranzberg's Laws"' (1986) 27(3) *Technology and Culture* 544.

⁸ Hildebrandt (n 2).

⁹ Andrzej Porębski, '27: Machine Learning and Law' in *Research Handbook on Law and Technology* (Edward Elgar Publishing, 2023) 450 <<https://www.elgaronline.com/view/book/9781803921327/chapter27.xml>>.

¹⁰ L Bennett Moses, 'Not a Single Singularity' in Simon Deakin and Christopher Markou (eds), *Is Law Computable? Critical Perspectives on Law and Artificial Intelligence* (Hart Publishing, 2020).

¹¹ Burkhard Schafer, '5: Formalising Law, or the Return of the Golem' in *Research Handbook on Law and Technology* (Edward Elgar Publishing, 2023) 59, 61.

¹² See Stuart Russell, Karine Perset and Marko Grobelnik, Updates to the OECD's definition of an AI system explained (29 November 2023), <https://oecd.ai/en/wonk/ai-system-definition-update>.

*An AI system is a machine-based system that ~~can~~, for a given set of human-defined **explicit or implicit** objectives, **infers, from the input it receives, how to generate outputs such as** ~~makes~~ predictions, **content**, recommendations, or decisions **that can** ~~influence~~ **physical** ~~real~~ or virtual environments. **Different AI systems are designed to operate with varying in their levels of autonomy and adaptiveness after deployment.***

In addition to introducing the concept of *inference* as a core element of the definition, it now captures the generation of content.

Back in the 1980s, lawyers interested in AI were focussed primarily on expert systems. This is, simplistically, a combination of a knowledge base and an inference engine. Legal experts without programming skills, working with a knowledge engineer, could formulate their knowledge of a domain (say, welfare law) into a computer-consumable format. The inference engine would allow the knowledge to be queried, for example to determine the payment to which an individual was entitled. While considered less cutting edge than systems based on machine learning methods, they had significant affordances. One advantage of this method is that reason-giving aligns with the logic of the system – a person could be told *why* they were not entitled to a particular benefit.

The ideas behind this are still relevant and are most neatly captured today in the idea of ‘rules as code’. There, the idea is to write (typically statutory, non-discretionary) rules in a machine-consumable format in such a way as to be as close as possible to the original natural language rule. In other words, looking at the ‘natural language’ and ‘code’ versions of the rules, the goal is isomorphism and careful cross-referencing. While perfect isomorphism is often impossible in practical applications,¹³ co-drafting and transparency offer many advantages from the standpoint of accountability.¹⁴ Building systems that implement the law using rules as code is certainly an improvement from subcontracted, non-transparent, unaccountable programs such as Robodebt.

The limitations of such explicit programming, where every possibility needs to be thought through at the outset and where update is manual, is answered by the idea of machine learning. This has facilitated, for example, data-driven predictions of litigation outcomes as well as improved the efficiency of tasks such as discovery. Essentially, and simplifying again, machine learning allows systems to identify complex patterns in data and use this to make predictions about new circumstances. Such predictions are based on correlations and patterns in the data rather than a process of legal reasoning – a system such as Lex Machina will predict the outcome of patent litigation not because it reasons through questions of novelty and inventive step, but because it knows that (say) a particular judge is likely to find that a pharmaceutical patent is valid when particular facts are present. The ‘reasons’ for a system’s prediction are thus different from the reasons one would expect from a judge. The system’s outputs thus do not make for a good argument in court.

One of the more exciting developments that lawyers have been watching is large language models. Large language models are able to generate text, so can respond to queries or create a legal document. They can also be combined with other functionality – so some tools involve a large language

¹³ Denis Merigoux, Marie Alauzen and Lilya Slimani, ‘Rules, Computation and Politics: Scrutinizing Unnoticed Programming Choices in French Housing Benefits’ (2023) 2(1) *Cross-disciplinary Research in Computational Law* <<https://journalcrcl.org/crcl/article/view/32>>; Mark Burdon et al, ‘From Rules as Code to Mindset Strategies and Aligned Interpretative Approaches’ (2023) 2(1) *Cross-disciplinary Research in Computational Law* <<https://journalcrcl.org/crcl/article/view/33>>.

¹⁴ 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 Opportunities for Public Law* (Federation Press, 2021).

model alongside, for example, access to a calculator, a library of documents, or the Internet. First, though, I am going to explain, at a very high level, how a large language model works and why they have become so much more impressive recently.

At its simplest, a large language model involves one file containing a lot of parameters and a relatively short 'run' file. The challenge is generating the parameters. That is typically done in two stages. The first is generation of the base model, which is created through training using a LOT of text (for example, the public Internet). That step is expensive and tends at the moment to be done only occasionally (as in the change from GPT-3 to GPT-4). The second step involves fine-tuning on a smaller set of documents, for example sample chats or legal documents. For example, ex Microsoft employee Peter Cole has trained a chatbot on Australian court cases.¹⁵ This step is not necessarily easy, particularly where the data being used to fine-tune the model is human-curated, but it is less computationally expensive.

Language models have used a variety of different machine learning techniques over time, with the general goal of producing the "next" word in a sequence. Well before ChatGPT became famous, we had Google search autocomplete. The technological leap that led to current levels of performance, which take us from writing a word or phrase to a paragraph, is the transformer architecture. As illustrated in a paper called "Attention Is All You Need",¹⁶ the architecture allows words to be understood from their context. Essentially, the process involves the following components:

- **Embedding:** Rather than processing words, computers process "tokens" (for simplicity, I assume that each token represents a word, although that is not necessarily the case). Each token is given a vector, essentially a series of numbers or you can think of it as a location in a multi-dimensional space. There are better and worse embeddings; a good embedding is one where words that are related are located close together and where it is easy to disambiguate words that can have multiple meanings. One can think of each number in the series or each dimension of the multi-dimensional space as representing a feature of the word (something that it will have common with some other words), although these may not be interpretable by humans. But ultimately when a large language model is working with words, it is in fact processing vectors.
- **Positional encoding:** The order in which words appear is important to the way that language works. Thus, when text is input into a large language model, the vector for that word must include its positional encoding, that is its place in the sentence.
- **Transformer blocks.** Like much machine learning, this is an iterative process, so there will generally be multiple transformer blocks. Each block consists of attention and feed forward mechanisms. The latter is essentially a neural network of many layers. The former is what has led to the recent performance leap. Attention uses context to disambiguate the meaning of each word, by assigning it to a new vector based on its context (being the other words). So, for example, imagine an input sentence includes the word "matter". That word will have an initial vector (as will every other word in the sentence). If the sentence includes words such as "court", "case", "litigation" or "law" – the vector of the word "matter" will change to move closer to those words. The other words in the sentence will also move, depending on *their* relationship with each word in the sentence. If, on the other hand, the sentence includes

¹⁵ Peter Cole, 'Courtaid.Ai: Chat GTP for Law' (25 March 2024) <<https://courtaid.ai/>>.

¹⁶ Ashish Vaswani et al, 'Attention Is All You Need' arXiv:1706.03762 [cs.CL]

words such as “atoms”, “science” or “elements”, then the vector for “matter” would shift in different direction. The model is thus able to disambiguate the meaning of the word “matter” in the sentence. The same principle disambiguates pronouns.

- **Softmax:** Ultimately, the output from the system should be a “next word”, because the large language model generates text one word at a time. However, a large language model is less useful if it is deterministic. That is because a deterministic model would output *exactly* the same story every time it was asked to compose a story. This step essentially ensures that the next word is sampled based on the probabilities of which is the best next word. It can be dialled up or down.

After this process is complete, there is a single word. That word becomes the next word in the output. It then adds to the string in the input that tells the system the “story so far”, and another word is generated. And so on.

Hopefully this very basic understanding of large language models is sufficient to understand some crucial things about the affordances and limitations of this tool in the context of legal work. Ultimately, it is a probabilistic text generator. It calculates the “next word” one at a time (like autocomplete) but relies on both positional encoding and the attention mechanism to track context. It does not know things the way that humans do – it does not, for example, memorise text (although its operation depends on its input text, which can include documents or webpages). It does not do analysis in the same way that humans do, so in what has been described as the “reversal curse”, a large language model may be able to answer the question “Who is Tom Cruise’s mother?” but not “Who is Mary Lee Pfeiffer’s son?”.

There are other important limitations. Large language models do not have a truth filter, for at least two reasons. First, if you ask an LLM to complete the sentence “The cat sat on the ...”, it will (mostly) answer “mat”. It will do so because that sentence appears many times in its training data, not because real world cats sit on mats. LLMs can also hallucinate, and many will now be familiar with examples. One that brought the issue home to lawyers was the example of a New York attorney filing a submission that referenced cases (that looked like cases with references in the correct format) that did not exist.¹⁷ The lawyer’s explanation that they asked ChatGPT whether the cases were real (the response was positive) does not help. Where accuracy matters, we still need to rely on humans.

There are also a variety of security challenges associated with LLMs. In addition to the need to ensure that confidentiality in input text is retained, there are problems with names such as prompt injection, jailbreaks, and data poisoning.

Present technology

The technologies through which law is created, recorded, transmitted, interpreted and understood *matter*. If law is carvings in stone, it is static and only able to be read by those who are literate and can

¹⁷ Mata v Avianca Inc (SDNY, 22-cv-1461 (PKC), 22 June 2023) (‘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.bbc.com/news/world-us-canada-65735769](https://www.abajournal.com/web/article/lawyers-who-doubled-down-and-defended-chatgpts-fake-cases-must-pay-5k-judgesays#:~:text=A%20federal%20judge%20in%20New%20York%20City%20has%20ordered%20two,then%20standing%20by%20the%20research; Kathryn Armstrong, ‘ChatGPT: US lawyer admits using AI for case research’, BBC (online, 27 May 2023) <; 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>>.

be in its presence. The law of the printing press is available to everyone who can buy the books. The law of the Internet era promises enhanced access, but also comes with new integrity challenges and a need for better information management. Further, it is not clear that access to 'more' in terms of volume makes for a better or more coherent legal system. If the medium of communication (oral, written, typed) changes how humans think in subtle ways, this will also influence how lawyers reason.

Law through the lens of generative AI is something else again. It is parameters applied to words, syllables and phrases, recombined through patterns of language both generally and in legal databases, and selected probabilistically to construct new texts. That does not make it useless, quite the contrary. Autocomplete can helpfully finish a letter ("I refer to [our earlier correspondence]"), create a factual chronology from documents and witness statements, or extract details for a procedural summary of a matter in a judgment. These are useful things. But we need to think carefully about how such tools might change what we are doing as lawyers.

Sometimes the use of technology misses the point. Leaving law for the moment, consider a common task assigned to high school history students. "Write an essay outlining the causes of World War I". Generative AI will perform well at this task – the essay will be produced with exactly the required number of words. But the use of ChatGPT to write the essay defeats the purpose of the exercise. The teacher does not actually want the artefact (an essay about the causes of World War I). Teachers ask students to perform this exercise because they want their students to understand that history is complicated, that there is no single explanation for what happened, that the reasons soldiers are sent to fight are often more similar on both sides of a war than those sides would be willing to admit at the time. They also want to help students improve the structure of their writing and their thinking, their ability to communicate complex ideas, and their ability to provide sources for asserted facts. If the student generates an answer using generative AI, there is no efficiency – indeed it is a waste of everyone's time.

Returning to LLMs as a legal technology, there are circumstances in which the use of AI is analogously pointless. Consider using a large language model to prepare an affidavit for a witness of fact. Such a document should express, in the witness' own words, their own personal knowledge of what occurred. Any use of an LLM is either putting words in the witness' mouth based on a miscellany of what they are likely to say or reformulating their words so that they sound more like an LLM than the witness. That might make the resulting document easier to read, particularly if a witness has a poor grasp of English grammar, but that isn't the purpose of the exercise. The resulting artefact will not be the witness' recollection in their own words.

The usefulness of generative AI tools in any context, including legal ones, depends on the alignment of the purpose of the task and the affordances and limitations of the tool. Preparation of a chronology from a large set of documents is fine and is a task for which today's computers are well suited. Such a chronology can be helpful to witnesses who can use it to remind themselves of the timing of particular events (such as when they sent an email). But it stops short of the problematic idea of using generative AI to literally put words into a witness' mouth. Like the history essay example, such a use defeats the point.

The use of LLMs to write a judgment faces similar boundary issues. Using generative AI to prepare a summary of the procedural steps taken to date based off documents in a court file is fine

provided there is no data leakage.¹⁸ Using AI to do legal research, as at least one Columbian judge has admitted to doing, will soon become inescapable as standard legal research tools increasingly move to natural language interfaces that rely on AI.¹⁹ What is important in legal research is not the technology as such but knowing how to use it. AI tools will require different skills from those currently used such as Boolean searching.

But what about writing a judgment? This is deeply problematic, as it undermines many functions of the exercise. If one thinks about Professor McIntyre's formulation of the judicial function in reaching a decision in a particular matter, I do not believe that a LLM can provide, for example, a "humanising touch to the abstract law", a difficult exercise of balancing stability and fidelity with change aligned to individual justice and social values.²⁰ No LLM, had they existed in the early 1990s, would have output a groundbreaking judgment such as *Mabo v Queensland*. LLMs reshuffle the existing corpus of concepts, words and phrases – they do not have the "spark of vitality that enlivens law within any society".²¹ We should not trust LLMs to exercise wisdom in making evaluative choices within the limits of the law²² – they have no meaningful life experience or familiarity with social values and expectations. Even if a reader believes an output is wise in a particular context, that merely reflects a lucky amalgamation of words that may have been associated by earlier exercises of wisdom. That is quite different from an exercise of wise judgment in the context of a particular case.

The problem is not only the lack of the hard-to-grasp concept of wisdom, it is a difference of method. Legal reasoning works with particular sources, and finding them is something that AI is very good at. But it uses those sources in particular ways – there are rules for statutory interpretation, ways to decide which facts are material, analogical reasoning that can be more or less persuasive, and so forth. Even if fine-tuned on legal texts, that is not what generative AI is doing. A large language model may *seem* to do some of these things, but its logic is not to implement a mode of reasoning, but rather to simulate it by drawing on patterns in samples of such reasoning, alongside text in input prompts. It might *look* like legal reasoning but it is not constructed through the methods of legal reasoning.

Does it make any difference if the LLM is given an extended prompt that sets out the judge's view on the case (based on the judge's own wisdom)? To my mind, such a question misunderstands the usefulness of the process of judgment-writing.²³ Judges often use the writing process to think through their reasoning. Some judges have, for example, reported not knowing the answer until the judgment is written or changing their minds in the course of preparing reasons. I raised before the issue of habits of thinking changing when moving from handwriting to typing, but surely moving from that to merely editing computer-generated text is far worse. This change would undermine the

¹⁸ 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>>.

¹⁹ '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-usesai-tool-chatgpt-in-court-ruling/>>.

²⁰ Joe McIntyre, *The Judicial Function: Fundamental Principles of Contemporary Judging* (Springer Nature, 2019) 69–71 ('*The Judicial Function*').

²¹ *Ibid* 71.

²² *Ibid* 94.

²³ Frank Kitto, 'Why Write Judgments?' (1992) 66(12) *Australian Law Journal* 787, 791–797.

“genuine reflection and deliberation” that contributes to the integrity and quality of judicial decision-making.²⁴

Further, if written reasons explain a judge’s reasoning for the benefit of the parties and the public, then the output of an LLM might be a believable proxy, but it will not actually be the output of a process in the judge’s mind. In that case, the reasons given would not in my view meet the requirements of open justice, which is a significant part of their purpose.²⁵

As a result, I disagree with those who suggest that the appropriate mechanism to decide whether an AI system can replace a human judge is to apply a Turing Test.²⁶ Under that test, the issue is whether a human can distinguish the output of a LLM from the judgment written by a human judge. The idea is that an AI system would ‘pass’ the test if a human is unable to identify a human from a machine output when not informed of its source. Some scholars have suggested this as a test for when AI systems might replace judges. For the reasons explained above, I do not think that the Turing test provides the relevant measure. The first challenge is one of timing – what seems wise or unwise immediately after a judgment may differ from what is understood much later. The fact that *Mabo* was much-criticised in its early days does not change the fact that our country is a better one for having overturned the doctrine of *terra nullius*. The wisdom of judges is best assessed over the longer term. But the larger issue is one of process. *Even if* probabilistically driven text generation reaches a level that experts find it hard to distinguish from the real thing, the fact that it wasn’t written by a judge matters. In other words, my argument is that justice is not only an output, but also a process. Litigants have a right to expect that a judge has applied their mind to the matter before the court and has engaged in a process that leads them to conclude that particular orders are appropriate. What matters is judgment, not the convincing simulation of it.

Another thing the Turing test fails to capture is legitimacy. Assuming one did replace human judges with an LLM, would one keep the public informed or actually invoke the veil of secrecy that the Turing test implies? Most ethical analyses of AI promote the importance of transparency, at the very least as to the fact that an AI system has been used.²⁷ European law also requires transparency in many circumstances.²⁸ Assuming people know that their matter is being determined by a probabilistic text generator, would they accept this? There may be differences based on both the court’s level, the court’s role and the issue before the court.²⁹ For example, an LLM authored default judgment or routine procedural orders will be less controversial than a disputed matter. But generally speaking, there is evidence that people tend to judge algorithmic decisions more harshly than human ones.³⁰ Alternatively, assuming people are not told, what impact on public confidence when it comes out? Either way, public confidence in the judiciary should be central, the risks of its loss are high.

Ultimately, where judges use AI, they need to be careful to ensure it does not take over what might be thought of as the core of the judicial function. In the language of Professor McIntyre, doing

²⁴ McIntyre (n 20) 274.

²⁵ Kitto (n 23); Spigelman, ‘Seen to Be Done : The Principle of Open Justice : Part 1’ (2000) 74(5) *Australian Law Journal* 290, 294.

²⁶ AM Turing, ‘Computing Machinery and Intelligence’ (1950) 59 *Mind* 433.

²⁷ Michael Guihot and Lyria Bennett Moses, *Artificial Intelligence, Robots, and the Law* (LexisNexis, 2020) ch 2.

²⁸ General Data Protection Regulation arts 13, 14, 22.

²⁹ Ray Worthy Campbell, ‘Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning Articles & Essays’ (2020) 18(2) *Colorado Technology Law Journal* 323, 334–341.

³⁰ Elina Treyger et al, *Assessing and Suing an Algorithm: Perceptions of Algorithmic Decisionmaking* (RAND Corporation, 2023) <https://www.rand.org/pubs/research_reports/RRA2100-1.html>.

so risks sacrificing a thousand years of institutional design for the sake of lesser goals such as efficiency.³¹

My point is not that AI should not be used by lawyers, or even that it should not be used by judges. My point is that it is essential that such use be appropriate in context. And deciding whether it is or is not appropriate requires an understanding of both the technology and the task. An AI system could, for example, be trusted with deciding disputes, but only with the fully informed and free consent of the parties – and only in the same sense that parties can decide to resolve their differences by tossing a coin.

Legal practitioners can use AI tools provided the human/machine process as a whole delivers accuracy and quality in line with professional obligations. There are still some systemic concerns if this were to happen at scale. We would, for example, move from bargaining in the shadow of the law to bargaining in the shadow of data.³² If lawyers cease to think like lawyers but rather observe patterns like data scientists and simulate reasoning like large language models, the law *in day to day practice* becomes something different. It might also be easier for bad actors to manipulate our decision-making using techniques such as data poisoning.

Bargaining in the shadow of big data can also entrench discrimination. I will give an example from family law. A number of years ago, at a family law conference, there was discussion of a tool to encourage property settlements that would give parties an indication of their likely share in a property settlement should the matter proceed. There are a number of fundamental questions that need to be asked about such a system including how gender, and in particular the different historical work patterns of men and women, is treated in the analysis. When I inquired about this, the answer was that these questions would be resolved further down the track. Worrying. But the problem isn't only bias one way or the other, but also the fact that this is now something we can control for. Should we? And who decides who decides – the project manager, the software developer, a data scientist, the court?

These are not easy questions. We can look to some guiding frameworks. Human rights is one critical guiding framework, as the Human Rights Commission has observed.³³ Use in courts requires attention to judicial values, including those we draw on in our work with the AIJA on AI decision-making in courts and tribunals, being open justice, judicial accountability and independence, impartiality and equality before the law, procedural fairness, access to justice, and efficiency.³⁴ Where law is involved and in the context of government decision-making, rule of law values such as accountability, transparency, consistency, predictability, and equality should all be considered in system design, even if we cannot guarantee 'rule of law by design' as such.³⁵ Transparency around the *fact* that AI systems are involved in a more than superficial or obvious way (say, grammar checking or internet searching) seems critical for *all* of these contexts, and some courts have already issued guidelines requiring this

³¹ McIntyre (n 20) 297.

³² Dru Stevenson and Nicholas J Wagoner, 'Bargaining in the Shadow of Big Data' (2016) 67(4) *Florida Law Review* 1337.

³³ Túlio Felipe Xavier Januário, 'Artificial Intelligence in Criminal Proceedings: Human Rights at Risk?' (2023) 21 *INACIPE Revista Mexicana de Ciencias Penales* 85.

³⁴ Lyria Bennett Moses et al, *AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators* (The Australasian Institute of Judicial Administration, 2023) <<https://aija.org.au/publications/ai-decision-making-and-the-courts-a-guide-for-judges-tribunal-members-and-court-administrators-2023-update/>>.

³⁵ Monika Zalnierute, Lyria Bennett Moses and George Williams, 'The Rule of Law "By Design"?' (2021) 95(5) *Tulane Law Review* 1063.

in the context of documents filed in the course of proceedings.³⁶ Guidelines can also have an educative function, reminding lawyers of risks associated with confidentiality and hallucination.³⁷

There is a lot of discussion about whether we need more rules and guidelines around current use of AI technology in critical contexts. I am generally against imposing technology-specific rules, particularly around a poorly defined concept such as artificial intelligence, because they quickly obsolesce. That is a different conversation.³⁸ But in my view, and you can see this in the AIJA report, the better approach is understanding combined with asking broad values-aligned questions.

Future imaginaries

Sometimes, when discussing the future of artificial intelligence in law, people opine that all of the current limitations of technology will be solved. So far, I have focussed on the limitations of existing tools but is there a future kind of AI tool that we *could* trust to run our legal system? Or are there aspects of the justice system that ought to involve humans *no matter what* kind of AI would be involved? People are, for example, looking into ways of making AI systems more transparent and large language models less prone to hallucination. Is it just a question of waiting for all the problems to be fixed, or is there some deeper reason for relying on human judgment, at least for some tasks?

I am not arguing that biological brains, by virtue of being made up of organic matter, are necessarily better than what we might be able to engineer in future. But everything depends on the detail. As we do not yet know that detail, I am going to turn to science fiction.

Imagine if human minds could be uploaded into a machine, could a person who is currently a human judge decide cases after such a ‘transfer’? There are some obvious limitations to such a ‘judge’. They would not age so could continue to operate indefinitely despite the fact that their sense of the community and its values are likely to continue to be tied to the generation with which they originally lived. We do not know, but the uploading process could alter who they are so that even if their memory retention were perfect, and they could continue absorb new knowledge as the law evolves, they may no longer have the same capacity for human empathy and understanding years. Or else it could decay years after their natural life.

Would we be comfortable if a judge used a brain computer interface, an idea made famous by Neuralink, while in court and writing judgments? In that case, I think we would want to ask about the process of thought and how it differs from that familiar to us. A human with access to and control over stored data or robot limbs raises fewer questions than the possibility that thought could be controlled or influenced externally, possibly without the human’s awareness. Other questions would relate to cyber security – to what extent could thought processes, or even data being drawn on directly, be ‘hacked’ by malicious actors?

³⁶ Dubai International Financial Centre Courts, Practical Guidance Note No. 2 of 2023 Guidelines on the use of large language models and generative AI in proceedings before the DIFC Courts, DIFC (December 21, 2023) <https://www.difccourts.ae/rules-decisions/practice-directions/practical-guidance-note-no-2-2023-guidelines-use-large-language-models-and-generative-ai-proceedings-difc-courts>

³⁷ Courts and Tribunals Judiciary (UK), *Artificial Intelligence (AI): Guidance for Judicial Office Holders* (12 December 2023); Courts of New Zealand, *Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals* (2023) <<https://www.courtsofnz.govt.nz/going-to-court/practice-directions/practice-guidelines/all-benches/guidelines-for-use-of-generative-artificial-intelligence-in-courts-and-tribunals/>>.

³⁸ Lyria Bennett Moses, ‘Regulating in the Face of Sociotechnical Change’ in Roger Brownsword, Eloise Scotford and Karen Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology* (Oxford University Press, 2017) 573; Lyria Bennett Moses, ‘How to Think about Law, Regulation and Technology: Problems with “Technology” as a Regulatory Target’ (2013) 5(1) *Law, Innovation and Technology* 1.

What about a legally expert equivalent of a character such as Data in Star Trek? For those who are not familiar with the franchise, Data is a self-aware robot with large knowledge databases and futuristic processing capabilities. His character is essentially a powerful robot, but at some point acquires an “emotion chip” that, in addition to being involved in various plot points, helps him become more human and empathetic throughout the series. Assuming Data could be programmed with not only legal knowledge but also the techniques of legal reasoning, would we accept Data as a judge? To what extent does public confidence in the judiciary relate to the fact that judgments are made by a fellow human, in the words of the Honourable Michael Kirby, “struggling conscientiously”?³⁹

This hypothetical eliminates some of the concerns we have with existing technology. We can assume (for the sake of argument) that issues such as hallucinations and bias have been satisfactorily resolved. At least in theory, having acquired human understanding through a (purely hypothetical) chip, Data may be able to navigate the complex tensions between integrity to the law and justice in the circumstance. The emotion chip may give Data the capacity for emotional intelligence that judges rely on when contemplating the impact of a crime on a victim, the motivations of an accused and the challenges experienced by divorcing couples.⁴⁰ However, it can also cause unanticipated errors and problems. Humans come naturally calibrated, as it were (for better or worse), but if “emotion” really is a dial that we can turn, where should it lie? And who decides who decides? Does having that power of choice make us more or less comfortable with the use of technology in sensitive contexts such as judging?

The hypotheticals are not, or at least not yet, important in themselves. But they highlight the need to understand the technology, in particular its affordances and limitations, alongside the context of its application. It is not, in other words, a question of projecting graphs past a point of “singularity” and then assuming that anything goes.

Conclusion

The use of AI in legal practice and the legal system, like many new technologies, can lead to dualistic utopian/dystopian thinking. The technology will always have enthusiasts, including judges keen to leverage capabilities and law firms embracing large language models into diverse workflows. Others will be far more cautious or simply lack the resources to put in place the security measures and training time that are a critical pre-requisite for adoption.

As mentioned, the AIJA report revolves around questions that we believe courts and tribunals should ask when deciding on AI use cases and the requirements that should be put on systems prior to adoption. I believe being able to ask the right questions is currently the most important skill for practitioners looking to leverage these tools. Law schools are increasingly looking to teach students how to ask the right questions about technology, which includes teaching them some things about technology, particularly tools such as large language models where our intuition about what they can do often deceives us. However, the people making decisions now – in government, in courts, in law firms, at the bar, in house – also need to acquire (if they haven’t already) a broad understanding of the affordances and limitations of the tools they are contemplating. They don’t need to become engineers,

³⁹ Tania Sourdin and Richard Cornes, ‘Do Judges Need to Be Human? The Implications of Technology for Responsive Judging’ in Tania Sourdin and Archie Zariski (eds), *The Responsive Judge : International Perspectives* (Springer Singapore, 2018) 98.

⁴⁰ For a discussion of the usefulness of emotion in reasoning about criminal justice, see Nina Peršak, ‘Automated Justice and Its Limits: Irreplaceable Human(e) Dimensions of Criminal Justice’ in Gert Vermeulen, Nina Peršak and Nicola Recchia (eds), *Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice* (Malku, 2021) 225, 229–234. See also Sourdin and Cornes (n 39) 97.

but they do need sufficient understanding to know what questions to ask and how to interpret the answers.

Our legal system is an essential component of having a society in which we all want to live. There are areas for improvement, and AI can enhance open justice through tools such as instant translation as well as expand access to justice by creating efficiencies. If we don't want to accidentally slip into some of the more dystopic visions, we need to keep abreast of technological developments and continue to ask the important questions. We should always remain conscious that the choices we make are not neutral but rather shape our legal system and how it operates for the people it serves.