Journal & Proceedings of the Royal Society of New South Wales, vol. 159, part 1, 2026, pp. 68–82. ISSN 0035-9173/26/01068-15

2025 Royal Society of New South Wales and Learned Academies Forum: "AI: The Hope and the Hype"

Panel Session 1: AI and the Law[1]

Emeritus Professor Rosalind Croucher AM FAAL FRSN FACLM(Hon), NSW Information and Privacy Commission

The Hon Dr Annabelle Bennett AC SC FAAL FAA, retired judge, Federal Court of Australia

The Hon Justice Melissa Perry, Federal Court of Australia

Professor Lyria Bennett Moses FRSN FASS, Faculty of Law and Justice, UNSW Sydney

Rosalind Croucher

I'm delighted to be able to chair this session on AI and the law. I thank you, Your Excellency, once again for hosting Ideas at the House and the Learned Academies Forum. This is such an expression of your own commitment to the work of the academies, the reach to community, and essentially all that we hope to achieve and aspire to do in the Royal Society. In introducing a panel on AI and the Law, a very necessary first panel if I may say in today's proceedings, I want to observe regarding Sally's framing of the day in terms of being "at the crossroads of rampant hope and rampant hype," that lawyers are not prone to rampant hype, and we keep our rampant hope for private conversations!

But it does provide a perfect prelude to a sage and sensible session on AI and the Law. And in embarking upon our conversation, Pascal, I have to be a little concerned with one thing that you said about AI and lawyers — that it would make lawyers "more productive." Usually words such as "productive," "productivity," and often their associated KPIs, send a chill down the spine of most lawyers, particularly judicial officers of whom I have two on the panel, but it can be a more finely calibrated assessment of what that "more productive" means. But I do bring it to the table as part of our thinking.

The panel that I'm chairing represents Fellows of the Australian Academy of Law. I'm delighted that the AAL takes its place among other learned academies in this forum today. As a quick background, the Academy of Law is the newest sibling, if you like, in the community of Learned Academies. It was launched in July 2007 at Government House in Queensland, when Her Excellency Quentin Bryce was then the Governor of Queensland. And it was the outcome of a process begun by the Australian Law Reform Commission in a landmark report on Managing Justice, a Review of the Federal Civil Justice System. The idea for the AAL was a key recommendation to bring together the three strands of the legal profession — the judiciary, legal practitioners, and legal academics — united in promoting high standards of learning and conduct and appropriate collegiality across the profession. (I note productivity isn't in there.)

The Academy of Law already holds an annual symposium with the Academy of Science, and today we expand that connection with other learned academies. The panel includes three extraordinary AAL

1 This is a transcript of a presentation, available at https://www.youtube.com/watch?v=6fEN3zWI-NM

fellows of great distinction — a tiny insight into which you can see in their cameo CVs that are contained in the programme. This panel exemplifies the eminence of those who are elected as fellows and the nature of the intersections of excellence across the branches of the wider legal community. The Honourable Dr Annabelle Bennett, AC SC, is also a fellow of the Academy of Science; next, the Honourable Justice Melissa Perry is a foundation fellow of the Academy of Law; Professor Lyria Bennett Moses, who is also a fellow of the Royal Society and of the Academy of Social Sciences. You are right to be in awe — as am I.

AI and the Law, the framing of the critical question for this forum sets the scene perfectly for our panel, namely: is AI truly different from previous transformative technologies, like electricity or the internet, or is it simply another innovation riding waves of hype? In 1970, there was a critical case in the High Court, *Mount Isa Mines Limited v Pusey*, in which Justice Windeyer spoke then of the law "marching with medicine, though in the rear and limping a little." The same can be said of any new technologies. In the rear? Limping? I pose in a rhetorical questioning way.

Each of the panellists will provide a short presentation. Dr Bennett will be looking at the tools we have now, particularly in judicial decision-making. Professor Bennett Moses will take us into a world when the law, or common law in particular, is not enough — and the tools of regulation. And Justice Perry will complete our wonderful triptych, in defence of the human. Throughout that, I'll add a few interrogative questions of my eminent team and then provide an opportunity for you similarly to have that opportunity at the end.

May I please invite Dr Annabelle Bennett, my very esteemed co-panellist to make her presentation.

Annabelle Bennett

Thanks, Ros. Yes. First, I have absolutely no expertise in AI, which is probably why I've gone first, but I do have part of my lifetime in the law. Before I start, I wanted to share with you an analogy that came to mind. Many years ago, there was a law conference in Alice Springs and we were all climbing Uluru, as was then permitted, which was then called Ayers Rock. And as I was coming down, the person walking in front of me was a rather large man. I took a photograph and obviously it was from behind him. And when that lovely man, and he was charming and became the Chief Justice of Western Australia, I sent the copy of the photo to him and I said, "What's this worth?," he replied, "It depends on whether you consider it a means unto an end or an end unto itself." And that came to my mind when I was thinking about this topic, because it really is a question of whether AI is the end unto itself, which would be replacing judges — for example, I am a sci-fi fanatic and used to watch Buck Rogers in the 25th century, Dr. Theopolis, who did everything and made all decisions and all judicial decisions — or whether it's a tool.

If it's a tool, then you've got to see how you use it. And of course, I'm picking up a lot of the other themes that have been mentioned because they are the common themes that come to mind. Let me just give you a couple of examples. When Dragon Dictate first came in, there was a question whether we could use that for draft judgments. It was a real question of editing because of course it didn't always transcribe things

perfectly. And if it, for example, missed out the word "not," because you spoke quickly as I do, then the answer that when you say the appellant succeeded or did not succeed, you can see that there can be a difference in the outcome of a case if you don't check it very carefully. That has happened to me once. I was walking into court and I was reading a little conclusion (this wasn't even AI) and I saw the word "not" had been left out. That had been through five edits and I only picked it up at the last minute. That's the sort of thing that can happen.

I would like to know how many of us have sent off a text message or an email where auto-correct has rather embarrassingly changed what you intended to say and something was totally wrong. So it becomes a question of editing and checking. I think it's a fantastic tool and we can use it, but again, we get down to two things.

Her Excellency raised the question of evidence. Of course, for lawyers and judges, evidence is paramount. You make decisions on the evidence that's presented, but then you have to be sure that the evidence is correct. And so you have to be able to check it out.

Let me tell you the trouble with checking — I'm picking up on your comment about provenance — sometimes you need to know what it is you're checking. Last week, I delivered the Sir Garfield Barwick Oration. Being a very, very good little girl, the first thing I did was to go to ChatGPT because I was going to speak about his extra-judicial activities. It gave me a wonderful list of his activities, and one of them said, with the years and everything included, "Chancellor of the University of Sydney." Now, I happen to know, because I've been in that academic world myself, both at the ANU and at Bond University, that he was chancellor at Macquarie University. So I typed back into my little conversation with ChatGPT, wasn't he the chancellor at Macquarie? And the response came, "Good pickup". But when I redid the search a few days later, just to check, it still told me the University of Sydney, so it hadn't changed.

But had I not known that, I would have delivered this lecture with all of his family present and made the fundamental mistake of saying he was the chancellor of the University of Sydney. I raise that as an example of a provenance issue. There's one other thing that I want to mention — and this gets back to the productivity — I feel like everyone's stolen all my thoughts in advance. There are two issues that come up here. First, if you rely on ChatGPT or AI to do your decision making, that is where the danger is. I would go so far as to say it's an impermissible delegation of the judicial power of the Commonwealth. There'd be enormous constitutional questions arising if you did that: whether you were actually exercising the judicial function. But at the same time, judges are actually very concerned with productivity.

And of course there's that longstanding expression, "justice delayed is justice denied." If you consider people hearing things without an electronic transcript, writing everything out by hand, thinking it through, rewriting drafts, then you can see the time it takes, even after a case is closed to get that decision to judgement. There's a lot of pressure on judges to get decisions out as quickly as possible. Tools have helped us do that: word processing, drag and dictate, all the other searches. Then you get into a balancing question of what's worth what.

What I've come to realise in a number of other areas that I've been involved in, especially well known in the commercial world, is the word "risk." You've got to have a risk matrix. You look at the risks and sometimes you have to accept that there is a degree of risk that is worth accepting in order to take advantage of what it is that a particular technology can offer. But that may mean that in particular cases, that risk will be realised. It's not much use to say to that litigant, "I made your decision because I checked that there were decisions around the world. I wanted to have harmony in the world, particularly in my area of intellectual property. I got a wrong summary from ChatGPT and I made your decision accordingly." It's okay to say, "Well, that only happens in 0.001% of the time." But for *that* litigant, that's the difference between success and failure, between prosperity and bankruptcy, between all of the consequences of that. I leave that with you.

The only other thing I thought I'd raise is interesting. I'm sure most of you would know that in the world of intellectual property, there are real questions to be asked about authorship and ownership. Bearing in mind that the whole idea of statutory rights, intellectual property rights, is meant to stimulate invention, to create an incentive for inventors to invent and to give their invention to the public, which is the quid pro quo for a patent. There have been cases all over the world now about whether a machine can be an inventor. Of course, some judges said yes, but ultimately most appellate courts have said no. And they've gone to the statute to say that the inventor still has to be a person. But the more interesting question has arisen, for example, in copyright.

There are two cases — one in China and one in the United States — that are most interesting, both of which, I want you to know, involved the creation of beautiful women. That's obviously something that engages a lot of people. Every female in this room would be the role model: all they had to do was take a photo and copy it. They'd be fine, but they went to the internet and took all these options that were available. I was told when I had to present on this at a conference in Singapore that there was a divergence between the United States and China on this. This goes to show, for anybody who's a lawyer or who's ever interested in the case, don't just read the summary. You've got to actually read the decision, because the decisions were identical in their reasoning, but the outcomes were different because of the way that technology was presented and described.

It's very important when you're presenting anything about technology — be it as an expert witness or in any legal case — how you describe and how you present it. Because in one case, the technology was presented as one that had human intervention at every choice, an actual decision being made by a human at each stage of the creation of this beautiful figure. And that was held to be a copyright work and that person was an author. But, in the other case, it was presented that it was really the machine that was making the decision, the machine-learning way, and you didn't have sufficient human intervention. And it was held then that there was no copyright work and there was no human authorship, so the person who was claiming authorship wasn't entitled to it.

I want you to know that, internationally in the legal world, these questions are

coming up. What is interesting is with the speed at which we're talking. Victor, this was part of the points you raised.

A few years ago we were talking about legal issues arising from the internet, simply the fact that something goes up on the internet and it goes all over the world, and where is the infringement taking place? Now it's getting down to a much harder question. It's not just theory now, because this is now impacting us, it is impacting the lawyers, it's impacting the judges, and it's an evaluation process. It's all very well to say you've got to check the provenance, but you need to have a sufficient degree of knowledge to check the provenance or have the tools available to do that.

It gets down to "what you don't know, you don't know." But I would say that I think it should be looked at as a tool for the judiciary and for the lawyers, but you've got to be very careful how you use it and how you check it and how you present it. But it has the possibility of increasing productivity, may I say, Ros, both for lawyers, because don't forget everything you can speed up — bearing in mind most lawyers charge by the hour, you will reduce the ultimate cost to the client, to the community — you'll hopefully get decisions out faster.

Hopefully they are just as good. Just as you haven't been able to do away with radiologists, I'm firmly of the view that you will not be able to do away with judges. I'm hanging onto that thought for grim death, even though I'm no longer a judge.

Rosalind Croucher: Thanks, Annabelle. The productivity question is an interesting one because it's about the use of tools and the distinction between a tool and the decider. I think that was very well put. I'm going to proceed straight into Lyria's presentation, which takes us into the world of regulation and where we can add things. Some of the conversation so far has been focused on the common law and the interpretive role of judges and decision making. We're going to broaden it out and we'll loop back on a number of themes as we go. So over to you.

Lyria Bennett Moses

I'm going to be relatively quick. One can talk about this stuff for a very long time, but I wanted to focus on the regulatory question. I noticed in the introductions a number of people brought it up. The idea of there's going to be all these issues and regulation can be one of the solutions. I want to unpack that a little bit, but what people mean by that, what they think it can do and how it might be approached. I was going to start with the definition of AI, but our esteemed Governor has already done that for us. There are only two things I'd point out. One is that the definition has changed. This shows the difference in the OECD definition of AI between 2019 and 2023. So it has changed quite significantly. The other one is to point out that when people talk about hope and hype, this isn't necessarily the targeted defined thing about which they feel that hope and hype.

I think the hope and hype has been continuous since 2019, even though the definition has fundamentally changed. At the core of the new definition is the idea of inference. These are systems that infer *how*. That wasn't even in the definition in 2019, although the continuity in hype and hope is there. If you think about inference, I don't think it's *inference* that people are concerned about. There are many things and inference can be part of it, but I think if you try to describe the AI that people have hopes for,

concerns about, you'll end up with other concepts coming in as opposed to the way it's currently defined. We're dealing with an emergent phenomenon, in other words. This is not a clear thing that we clearly understand its current existence and future. But I also want to highlight that hope and hype is also something we can feel about regulation. I think there has been a lot of hope in regulation in some of the introductions — that this can be the thing that protects us. Let's see if that is in fact the case.

So why regulate? There's a whole range of issues that come up in AI. I picked one: the algorithmic bias or discrimination problem. Most people are familiar with at least some of the examples of this: these hiring algorithms that prefer men, et cetera, et cetera. The point is that systems, defined AI systems, when they discriminate, do it differently to humans. A human might discriminate say on the grounds of sex because they have a particular view of women as not being able to perform a particular job, whereas systems will typically be black boxed. It'll be probabilistically related based on historic data, influenced potentially by variables such as sex and race or what have you — even when it's not directly impacted by those variables, because those variables are not inputs into the system, impacting people differently by proxy variables based on similar characteristics. My point is that the challenge that we face around discrimination, around fairness, is not an AI problem. AI is part of a broader problem that is all in fact connected to each other, often through the data on which these systems are learning. So what does the law currently do? The core point is that the law, as it is currently written, was for the imagined scenario, which was at a time when the law was written for a human who might be sexist, racist, et cetera.

The goal of the law is to stop that person making a decision, say not to hire a woman because she's a woman or in a whole variety of other contexts. That is what the law is designed for. We have different laws in Australia. It's a bit of a complex mess. There's state law, there's federal law, there's different laws sometimes for sex discrimination and race discrimination. They all use slightly different words. This is not going to be that lecture, but to take direct discrimination as an example, you'll typically find words in the legislation like "on the grounds of." Does someone make a decision on the grounds of, say, sex? Of course, when you're looking at a *system*, it's not that clear. We can't talk about "on the grounds of" as if systems are doing it on the basis of one thing, or that we'll even *know* what the grounds are for a particular system output.

It's a lot more complicated. You have other concepts and terms, but where you often go when you're looking at the kinds of things that systems are doing in the context of unfair treatment is we move into *indirect* discrimination, which is where you say the condition of hiring is being able to go through some hiring algorithm which goes forward to the next stage and has the effect of disadvantaging people, say on the grounds of sex. We're going to pick different acts here, but that's the core concept. The problem of course is that this will almost always be the case, at least in some conceptions of what we mean by different treatment. In other words, it's very hard to think of a system where that will make absolutely no difference at all.

What happens then is the wonderful machine-learning community tries to

enlighten us. And they create a whole raft of definitions of fairness. There are over 30 of these things, and they're very precise because the one thing I love that comes from that community is its precision. When we talk about fairness in law, we talk about discrimination, we have a general sense of what that is. We use words, they use mathematics. It is a much clearer sense of what particular fairness metrics might require. That is fantastic except for two things. One is they're mutually inconsistent. You can't do all of them at once. And the second problem is that how does that fit in with what the legislation is telling companies to do in discrimination legislation? And it is not a great fit.

Statutes are built on words, as opposed to common law, which is a little bit more flexible. Statutes are built on words and those words get interpreted. "On the grounds of" means what "on the grounds of" means. You have this language and legislation built around one problem type, the sexist racist human. You have systems that can potentially solve for some of these problems. And then you have law that's not actually helping organisations to make sensible decisions and may in fact be pushing them to do things like eliminating variables that might actually increase the problem, but nevertheless avoid categories like "direct discrimination" that the legislation prioritises and has fewer exceptions. You've got a misfit. It's a complicated misfit, but you've got a misfit between what these systems are doing and the problems they're creating and what the law is requiring organisations to do.

Ultimately, you have machine learning trying to help, but the law's not talking to those kinds of distinctions and we're not necessarily getting what we want. So what do we do with regulation? Do we do something? And I think regulation often becomes a narrower conversation than it needs to be. This is my very giant picture of regulation. First of all, not all of it is government. Standards organisations are regulatory. There's a whole bunch of other stuff that goes on. But even within government, there is a whole bunch of stuff. It passes laws. It creates ethical principles and in particular has done so around AI. There are guidelines and all sorts of other documentation. It buys things or doesn't buy things based on particular criteria. It has a whole bunch of influence levers.

Let's even narrow down to law. The first point to make is that not all law that is regulatory is going to have the words "AI" in the heading. It's not all about the "AI Act." I've already spoken about discrimination law. Now that isn't a good fit, but it nevertheless influences decisions in organisations about the kinds of AI systems they build. They don't want to fall foul of those laws. We can talk about law being inadequate or not sufficient, but we can also think about the role those laws might continue to play if they are done better than they are now. Of course, you have the possibility of law that specifically targets AI. But there you have problems — because again, you've got to define it. Here's the EU AI Act definition:

> "AI system" is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Again, what you find is that the act targets a technology. It's one imagination of what we're scared of, but it's not necessarily always the right one. The example I like to use to show this is one of the prohibitions in the AI Act.

> The following artificial intelligence practices shall be prohibited:
>
> (a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm
>
> (*Artificial Intelligence Act* article 5(1)(a))

You can read it if you want, but I want you to imagine leading the text in red. These are about the use of subliminal techniques to manipulate people. We don't want that, is what they say. That's prohibited. But only when you use an AI system, only when you fall within this very particular conception of a particular technology and not in general. But why not? There's nothing in this where I feel that the removal of the red text would make it worse and it would in fact expand it to cover a broader range of the problem of manipulative techniques. That highlights, I think, some of that problem. And where do we end up? We have choices in regulation. We can focus on the technology, we can regulate AI, but if we do that, we're focusing on a current imagined definition of what we think this technology is and we're passing laws about that *particular thing* and imagining *that* is the problem.

We're not seeing it, as I started with in that broader context of broader problems where the technology may be playing part of a role, but isn't the whole story. It will become obsolete because whatever we think AI is now may not be the thing that's coming. There's a whole bunch of problems in doing it that way. One can do it the way the US did under the former Biden administration, which is sector-based. You do something in education, you do something in commerce, you do something in housing, et cetera, and you say, these are the rules for that sector. That's another approach. Or you can — and this is the one I like to push — start with the values. Start with what problem you're trying to solve for. If we say we say want to solve the problem of discrimination or certain kinds of discrimination, then you need to think quite deeply about what are your red lines, what are the kinds of practices and behaviours you want organisations to do when running systems, say for hiring people or anything else.

What are the tests we want them to run, the evaluations we want them to do? What *matters*? Because then we can look at all of those fairness metrics that the machine-learning community has very helpfully created and help guide organisations in making the right decisions. That is not necessarily a case of an AI Act. That's a case of looking back at all of those discrimination laws and thinking about what we really want.

Rosalind Croucher: Thank you, Lyria. The drawing attention I think always is on "what is the problem?" What is the problem we're

trying to solve? And then not doing that lightly because often law can sometimes or legislators can sometimes do things quickly, the rapid response — "we need to have a law on X." If it's not thought through properly, it can be compounding in its consequences — the unintended consequence of laws. I'll give you just one quick example before I invite Justice Perry to speak. In the mid-1980s, it was the period of burgeoning use of reproductive technologies, a wonderful assistance to families and to having children. There was encouragement to use assisted reproduction through donor sperm. There were laws passed out of concern to preserve the anonymity of the donor — and it was just simply that. The sperm donor is not the father except in very, very narrow circumstances.

The sperm donor was irrefutably presumed not to be the biological father; problem solved. But what that did was completely overlook the rights of the children who might be born as a result. Their curiosity, indeed, their rights to know their own biological and genetic history, which has led much later to responsive amendments that are in line with the changes that came in the adoption area, where children were able now to ask questions about their own biological past. What is the problem we're trying to solve? The next two parts of the formula are: what are the possible solutions, and then what ones should we adopt? Melissa, could I invite you to address us?

Melissa Perry

Thank you. In the movie, *Eye in the Sky*, senior military personnel from the British and US authorities were advised that it was lawful under international humanitarian law to order an attack from a drone on a cell about to undertake a suicide bombing. Collateral damage being the likely killing of a little girl, selling her mother's bread from a makeshift stall proximate to the attack was considered proportionate to the harm that would be caused if the suicide bombing were to proceed. However, the agent on the ground was moved by the little girl's plight. And with barely a moment to spare, he acted to save her by bribing a local girl to buy the last of her loaves in the hope that she would leave before the assault.

This example illustrates how an exercise of judgment by a human decision maker may mean achieving not merely a *lawful* result, but one which better accords with the fundamental human values of mercy, fairness and compassion in the individuals and the public interest. This is not a mechanical procedure, which is readily reduced to an algorithm leading to an inevitable result, even if that result is not predictable by a human. The example also illustrates the ability of humans to problem solve in creative and imaginative ways. Such fundamental human qualities as mercy, compassion and fairness draw upon our shared understanding of the frailty of the human condition, and lie at the core of our common humanity, resonating through the ages across cultures and underpinning fundamental human rights. Human qualities of the kind which I've described have long informed courts and government decision makers in the exercise of their functions.

While the width of statutory discretions will vary according to context, discretions potentially afford humans the latitude to make judgments and reach decisions, which reflect community, administrative, and international values, and align with statu-

tory objects in the face of a wide or almost infinite variety of human circumstances.

However, these are qualities utterly lacking in any machine technology. No robot has yet been created with the capacity for sentience, understanding, or independent thought, let alone with a conscience, which is a key, if not often a determinative, element used by humans to make choices and reach decisions. Thus, while generative AI can emulate the process of exercising a discretion and may even provide reasons which give the appearance that it's exercised a discretion or engaged in an evaluative process, this is merely smoke and mirrors. This is because the processes which it utilises are based upon statistical analyses and not upon the nuanced reasoning processes in which humans engage. For example, there are systems available which will assist in predicting the outcome of litigation, and these may operate with a high degree of accuracy and be infinitely more cost-effective than engaging a lawyer, but they may have regard to largely irrelevant factors, such as the name of the judge and the amount of money involved in the case.[2]

This leads me to a concern that the massive efficiencies and cost savings which the use of machine technologies potentially afford in government decision making, especially high-volume decision making, such as determining eligibility for social welfare benefits, may encourage governments to frame laws in such a way as to authorise the use of technologies to make the decisions. This in turn may curtail the scope for an exercise of judgment and discretion by a human decision maker to accommodate borderline or unanticipated cases. These concerns are heightened in the context of decisions which affect human rights and fundamental interests. Indeed, even the use of such programs to assist in the exercise of evaluative judgments and discretions needs to be approached with caution to ensure that the human decision-maker brings a properly independent and impartial mind to bear on the issues and interrogates the data provided by such technologies.

An example about which much has already been said is one form of generative AI being a large language model or LLM, such as ChatGPT, Copilot, and Grok. An LLM is a complex algorithm which responds to human prompts to generate new text, representing the most likely words and the most likely word order based on training from massive data sets. The text, therefore, is ultimately no more than probabilities, but the manner in which generative AI functions has a number of consequences, which are also highly pertinent when consideration is being given to the use of such models in the judicial and litigation spheres. I wish to highlight three short points.

First, as Lyria has already mentioned, the risks of hidden bias influencing outcomes from generative AI are based on the fact that the data on which it is trained is historical, but it goes deeper than this. Because generative AI operates by formulating the most likely sequence of words in the most likely order, it may limit exposure to a range of different and competing views, particularly minority opinions, and further marginalise minority and disadvantaged groups in response to prompts. Second, the outputs of LLMs are unpredictable. Devoid of understanding or concepts of accuracy, the

2 See Johns F and Walton K (2026) Why AI will redefine the law degree. *Sydney Morning Herald*, 21 April, p.21, for a discussion of the continuing value of legal education. [Ed.]

capacity of LLMs to hallucinate or make up information and to convey these hallucinations convincingly is well-documented and already been referred to. Infamous examples in the courtroom context include fake citations of earlier judicial decisions generated by AI on which parties have sought to rely to their peril. Indeed, somewhat ironically, in reporting on a recent report prepared for government by a professional services company using AI, the ABC 7.30 Report identified a number of hallucinations, including a fake citation to an article purportedly by me.

Added to this, the way in which outputs have been generated by LLMs are generally unknown and unknowable, even by their developers and programmers. In other words, the systems operate (as Lyria has explained) as black boxes, having completely opaque decision-making processes. Not only therefore is the reliability of output a significant issue, but the use of LLMs raises serious transparency, accountability, accuracy — and therefore audit and reliability — issues. But the metaphor of a black box is also apposite in another respect, given the inability of LLMs to think outside the box in the way that the example I started with illustrates.

The final point I would make is that it isn't just the outcome which matters, but also the *process* leading to the outcome, especially where the outcome is one directly affecting the individual. In other words, AI may give the same answer which a human decision maker would give, and it may give it more efficiently and more quickly. It may also provide the answer drawing upon a vast amount of knowledge well beyond the capacity of a single human being to master. But the process matters, and the fact that reasons given for the outcome are the product of a human being who listened and considered the evidence and the submissions, and brought to bear the human qualities of the kind I referred to at the start in reaching their decision, matters to those whose lives are affected by decisions which we as judicial officers, administrative decision makers, and decision makers in the private sector make *and* take responsibility for.

So those observations are not merely based on my experience as a judge of almost 13 years, where the importance of ensuring that litigants have a real and effective opportunity to be heard and to feel respected and listened to is brought home to me on a daily basis. Studies based upon interviews with litigants show that confidence in the system depends upon the litigant feeling that they were *heard* and not necessarily upon whether or not the particular litigant is successful.

Taking a practical example which I'll finish with. How might *you* as the victim of a scam in which you lost all your savings feel, if instead of a human judge, you read out your victim impact statement to the sentencing robot judge, would you feel listened to? Would it be *enough* to know that the robot judge was more likely to reach an appropriate sentence in line with precedence and community values than the flawed and fallible human judge with a living beating heart?

Rosalind Croucher: The difference in the process of reasoning: we will focus on the judicial decision making for the moment. The law historically is a large language model. Now we've got a lot of law, but that is also the beauty of the way that common law, which is the sum total of decisions that have been made by judicial officers over many

centuries now, is a large language model and is amenable to the process that large language models use. It's when you get to the finer points of the reasoning, and I think, Melissa, you illustrated that beautifully in terms of the human and the flexibility that Annabelle, you spoke of, where the judge can be innovative within the confines of flexible ideas like copyright and who is the maker and so on. There is room in those spaces for that reasoning that is deeply humanly, personally informed, but within the constraints of important processes.

I was thinking in anticipation of our conversation, that when judicial decision making is based on both the use of precedent, and the possibility of leaps when the precedent runs out. I'll direct this maybe to Annabelle, I think of a case like the Mabo decision, which was one dissenting judge and the remainder of the bench where it was a unanimous or unified decision, and it *shifted* the parameters of the common law in relation to the recognition of native title. That leap, how could that possibly happen if informed or otherwise by AI? Not the actual decision, but I can't imagine a world when that decision might have been made relying on the AI tools that are currently on offer.

Annabelle Bennett: That's a hell of a question, Ros. A couple of things that came out as you were talking, though. The first is, you spoke about the common law, and of course that is one of the great strengths in our system, that it gives the degree of flexibility, I think, very much in the human element. If you've got the common law, part of our history is you *can* develop it. What has happened is with a lot of dissenting opinions, over time, those dissenting opinions have become, as the majority have moved towards them, the majority of views. David, my husband, is here, and we acted for Lionel Murphy. At the time when he first came out with some of his decisions, people were appalled about how extreme they were, but of course, in a lot of cases, things have shifted and they're mainstream.

Coming back to the AI thing, the point that Lyria was making is that judges don't *always* have the flexibility to make the decision that they think should be made in the circumstances, because if there is a statute that covers *exactly* what you're looking at, you *must* give that statute effect. And this is where the difficulties are, the sorts of difficulties that Lyria spoke of, of statutes fixed in time. I'll give you one small example that's not in this area, but in terms of trying to do what you feel you want to do and being constrained by statute, I sat on the Court of Arbitration for Sport, I was at the Olympics, there was a wrestler, an Indian woman wrestler who got to the finals — first ever. She failed the weigh-in and the rules that I had to construe said, "up to 50 kilograms." She was a hundred grams over. "Up to 50 kilograms," and she was disqualified.

I wrestled with that all night. I actually wrote it to give her a win, and I woke up in the morning and I said, "I can't do that," because the words of the statute are so clear that, in this case, the regulation was so constraining.

The first point is that this is the flexibility of our system that you described so well. Melissa has given the example of the beauty of giving judges the flexibility to deal with the facts of a particular case. Where you come in though is another one, which there are hundreds of thousands, if not millions of decisions that one could go to to make a decision. A lot of them are dissenting deci-

sions, which no one usually talks about until you look for them later.

Would AI be a fabulous tool to be able with the right prompts to give you that information, which you would never find, and practically you would never do it? Yes, you could have it — as long as you were satisfied of the provenance of it, but you'd have to check that. You could get the little summaries and go away and read the decision and see if it was right for yourself. AI could be a tool to go across the world to look at decisions. Google Translate can tell you if it's from another language. And, yes, AI could be very useful, but, ultimately, as Melissa said, it is up to the individual judge then to take that information, check it through, think about it, take bits of whichever you think are good, and then synthesise it in your own decision-making process and come up with what you believe is the right answer in that case.

Rosalind Croucher: Thanks, Annabelle.

Lyria spoke of the dangers of shortsighted regulation. Annabelle was speaking of dangers, particularly in terms of AI judges — in other words, relating it to the constitutional framing and delegation of judicial power. The idea of judicial power is something not amenable to replacement, because it lacks the human, because it also lacks that ability to discriminate in the way that judicial officers spend their entire life training for, and all of the legal profession. But the shortsighted, knee-jerk regulatory responses, the dangers that might lie there and building on the other observations of the panellists — these are Annabelle's dangers.

Lyria Bennett Moses: Victor, you might have more first-hand experience of this, but I think that the challenge for politicians is that people are often scared of technologies. AI is a really good example of that. It's very nice as a politician to be able to announce, "We are going to do something about this thing that you're scared of. We are going to regulate AI," et cetera. It's far harder to say, "we are going to clarify some of the challenges of interpretation of discrimination law." We're going to have to look at a whole range of acts, because as I said, it's not in one act, and we're going to really think about how to align those values with what society cares about so that it can apply — whether it's a human, a system, or a hybrid decision. It's a lot more work, and you don't get the punch of saying, "We're doing something about the technology you're scared about." And you have to do it *over and over* because people aren't just scared about the fairness of AI systems. They're worried about accountability. They're worried about a whole range of things. You have to keep doing that. It's hard. Getting things right, or the best possible answer, is more work and less reward. Whereas if people are scared of something, if something bad happens and you can say, "We have a law that is going to constrain this beast that you're worried about — this AI," it's attractive.

Rosalind Croucher: We have room for two questions, and then I'll end with just a couple of sentences allowed to me as the chair.

Q1: In a criminal system, if an AI system commits a crime, how is the crime judged? Is it "we are going with the accountable human" and judge the *mens rea* and *actus reus*? And will there be also a human dignity issue: who faces the crime? How complex is that scenario?

Annabelle Bennett: How long is a piece of string? Lyria might have an opinion on that too. First, judges are used to evaluating

evidence and the weight of evidence. It's one of the things we do all the time. If you had a human giving evidence against some sort of machine output, you would have to evaluate that. In the ordinary course you would do that. I don't know about an AI committing a crime. This gets down to the question of who's responsible. It's a bit like, can you get an AI, can the machine be the inventor? You'd have to look at the statute and you'd have to see whether it's a person who's required. Then ultimately you get to who wrote the algorithms and who marketed it. It depends on the particular crime you're talking about.

But these are clearly questions that eventually could well come before the courts. And I don't know what else Lyria or Melissa might add.

Lyria Bennett Moses: I can have a go at it. So first of all, statutes are drafted as statutes are drafted. At the moment, most of the kinds of crimes you'd think of are basically crimes committed by people. At various times in history, people used to convict animals. We don't do that anymore. Crimes of statutes are written about humans. Should we expand it to systems? I think that is an interesting question, but also a confusing question because systems are not motivated by the same kinds of things that motivate humans. If you're trying to stop humans murdering people, one good way to do it is to say you will spend a really long time in jail if you do that. I don't think that's necessarily motivating for systems as such, or what jails for systems might look like, but we can think differently about how to stop systems doing bad things. And we have to sort of ask those questions rather than just assume current law and just add "or robot" or something like that in it.

Melissa Perry: The question that you raise is a very interesting one because it leads into the question of who is to take responsibility for the outputs from machines. Is it going to be the person who uses the machine to produce the output, or is it going to be the programmer, or is it going to be the person who has responsibility for those who have responsibility for auditing the system? So there probably does need to be some clarity around those sorts of issues. They're very fundamental questions. Second, in the administrative government decision-making context, we do have a lot of provisions now in statutes which raise some rather interesting constitutional issues, but purport to not only authorise the use of machines in government decision making, the exercise of statutory powers, but also to deem any aspect of the decision making process undertaken by the machine to be made by an officer of the Commonwealth, so by a human decision maker. In that way, the statute itself has expressly dealt with *who* is to take responsibility for the output from the machine or any flaws in its processes.

Annabelle Bennett: So waving a magic wand and saying, "I hereby deem this machine to be human."

Melissa Perry: Yes, effectively, and a specific human.

Rosalind Croucher: Thank you. I think we'll leave it at that question. We began today in a world of, I think, hope — the possibilities of AI as an enormously useful tool, the world of hype that can transform complex decision making over large data sets into seconds, decisions that are assisted by analysis produced in seconds. So that's enormous areas of hope. Technology can certainly fix, it can enhance, and it can transform. Where laws and legal frameworks come in is that law

and the possibilities of legal frameworks and regulation can legitimise, and it can also hold to account so long as it is anchored in the world of people, and a world of humans and a world of humans with human rights. With that, I would like you to thank and warmly congratulate my wonderful panel.