REPORT ON PROCEEDINGS BEFORE

PORTFOLIO COMMITTEE NO. 1 – PREMIER AND FINANCE

ARTIFICIAL INTELLIGENCE (AI) IN NEW SOUTH WALES

CORRECTED

At Macquarie Room, Parliament House, Sydney on Monday 11 March 2024

The Committee met at 9:30.

PRESENT

Mr Jeremy Buckingham (Chair)

The Hon. Robert Borsak (Deputy Chair)
Ms Abigail Boyd
The Hon. Dr Sarah Kaine
The Hon. Stephen Lawrence
The Hon. Jacqui Munro
The Hon. Cameron Murphy
The Hon. Chris Rath

The CHAIR: Good morning, everyone. Welcome to Monday morning and the second hearing of this Committee's inquiry into artificial intelligence in New South Wales. Firstly, I would like to acknowledge the Gadigal people of the Eora nation, the traditional custodians of the lands on which we are meeting today. I pay my respects to Elders past and present and celebrate the diversity of Aboriginal peoples and their ongoing cultures and connections to the lands and waters of New South Wales. I also acknowledge and pay my respects to any Aboriginal and Torres Strait Islander people joining us today.

I ask everyone in the room to please turn their mobile phones to silent. Parliamentary privilege applies to witnesses in relation to the evidence they give today. However, it does not apply to what witnesses say outside of the hearing. I urge witnesses to be careful about making comments to the media or to others after completing their evidence. In addition, the Legislative Council has adopted rules to provide procedural fairness for inquiry participants. I encourage Committee members and witnesses to be mindful of these procedures.

Professor ADAM BRIDGEMAN, Pro Vice-Chancellor Educational Innovation, University of Sydney, affirmed and examined

Professor DANNY LIU, Senior Academic Developer, University of Sydney, sworn and examined

Ms AMBER FLOHM, Deputy President, NSW Teachers Federation, affirmed and examined

The CHAIR: Welcome, and thank you for making the time to give evidence and to make a submission. Do any or all of you have an introductory statement to make?

AMBER FLOHM: I certainly do.

The CHAIR: We will start with you, Ms Flohm.

AMBER FLOHM: Thank you. I also would like to begin by acknowledging the Gadigal people, the traditional custodians of the land and, of course, our very first teachers on our homelands. I pay my respects to Elders past and present and extend that respect to Aboriginal peoples who may be present with us today. Thank you to the Committee for the opportunity to contribute today to the inquiry, to appear as a witness and, of course, for taking the time to read our submission. AI undoubtedly presents opportunities for public schools and other education settings in New South Wales, and society more broadly. However, there is an urgent need to face the risks and challenges of AI to ensure it is implemented in a safe, informed and equitable manner which respects the rights of students, teachers and parents. AI is likely to have far-reaching implications for the economy, democracy and society at large.

Focusing on the impacts of AI on public education, as the underpinning of our democracy, prosperity and a harmonious and fair society, is really central to the evidence that I'll give today. Getting AI wrong could well be disastrous for public schooling. The federation wholeheartedly agrees with this Committee's media statement, made on 28 June, that "failure to plan and to regulate is simply not an option". It's essential that AI, including generative AI, in New South Wales public education settings is implemented in genuinely responsible and ethical ways that demonstrably benefit our students, teachers, schools and society. Teaching and learning is inherently an intellectual, human and social practice. Teachers are innovative, adaptive and future-focused professionals, guided by their students' learning needs and their sense of dignity and pride in their work. They nurture the inherently social, human and professional relationships they share with their colleagues and students, including the exchange of knowledge.

It's imperative that humans, especially workers, direct the use of AI and always remain in control. Machines cannot think for humans, and for teachers, their professional judgement and independence must continue to drive all stages of the teaching and learning cycle. AI must be recognised for what it is—a tool to be used in accordance with a teacher's professional judgement for the purposes of enhancing educational outcomes. The professional judgement of teachers, their qualifications, their training, their expertise, their skills and their experience ensures that they are able to address these matters as they arise in their classrooms, that they are able to address through curriculum, professional standards, policies and the legislative frameworks in which we operate. This is the core.

Teachers, however, will require significant support to enhance the enabling conditions within the public education system. This goes to the matters of the current context in schools: continuing teacher shortages; underfunding; high levels of burnout; attraction and retention rates which are still unacceptable; and an unsustainable workload, which is volatile to changes in demands on the profession. Professional support and development for teachers in the area of AI will also remain critical to its successful use. There is also substantial evidence of AI's potential to cause social harm and to contribute to engineered inequality in educational context, as demonstrated through numerous examples of minority groups being systemically disadvantaged when AI models amplify discrimination baked into their training data.

It would be prudent for the New South Wales Government to keep in mind the findings of the important UNESCO report released last year addressing technology in education. They said there was little robust research to demonstrate digital technology inherently added value to education. UNESCO said in its *Global Education Monitoring Report 2023* much of the evidence was funded by private education companies trying to sell digital learning products and their growing influence on education policy around the world "is a cause for concern". Fully qualified teachers, as experts, must therefore retain decision-making power over questions of pedagogy and andragogy, curriculum implementation, teaching strategies and resources, assessment and reporting, and analysis of student learning data and outcomes. They must retain scope to use their professional judgement, specialist expertise and knowledge of students and how they learn to determine if, when and how digital technologies are used. They must be able to determine that they are fit for purpose and are truly benefiting our students.

Today, I proudly represent over 60,000 public education teachers across New South Wales. These are the workers that our communities across New South Wales rightly rely upon to secure the educational and wellbeing outcomes for our children, our young people. We must do everything to ensure that our students have futures which are happy, fulfilling and contribute to the economic prosperity and social cohesion of our society. We are proud of our public education system in New South Wales and its inclusive nature to meet the complex and diverse needs of our communities. Within that frame, Teachers Federation members will do as they have always done—put the learning and needs of their students first—and the adaption of AI in our public schools will see no diversion from this well-known path of commitment and dedication. Thank you.

The CHAIR: Thank you, Ms Flohm. We'll begin with questions from the Government members.

The Hon. STEPHEN LAWRENCE: I am curious to hear the perspective of each of the three witnesses, maybe starting with Ms Flohm, about the challenges that artificial intelligence poses to student assessment, maybe starting with what is going on now; are students using AI to cheat? What exactly is happening out there? Then maybe a bit of a projection into the future in terms of challenges posed.

AMBER FLOHM: In terms of number of students that are using AI for assessments, the Teachers Federation doesn't hold that data, but what we do know anecdotally is that particularly our secondary students are using AI, and generative AI mainly. In the area of assessment, as you'd know, ChatGPT is banned in public schools so the monitoring of that is not possible, although I am aware now that the Department of Education has the EduChat app currently being piloted across sixteen public schools for terms one and two to look exactly at that data and how students might use them. The greatest concern from the profession in relation to their students and assessment is really around data and privacy security, and making sure that teachers feel confident that the data that is being collected is safe and that the privacy and what we call the appropriate guardrails are in place to make sure none of that data is leaked to other sources. Unfortunately we saw during the COVID pandemic, when the profession flipped to learning from home, the use of those really big edu-businesses—Google et cetera—and how that student data was used way beyond educational opportunities and sold on. That is a legitimate concern around data and assessment.

DANNY LIU: I think the research coming out of the various places around the world is indicating that the propensity for students to use AI to cheat is probably not as bad as we thought it was going to be. At the University of Sydney our approach to assessment and AI is to acknowledge that students will have access to it even if we tell them they can't use it so we've basically said that in order to satisfy our requirements of confirming that our graduates know what they know, we need to have in the program some areas of highly secured assessment but for most other assessments we need to have the ability to help our students actually engage with AI in a productive and meaningful way.

We have what we call a two-lane approach to assessment which is saying that in a program we'll need some highly secured assessments—there might be sit-down exams or oral examinations where we know it's the student—and then, for the most part, we want to have assessments which are authentically engaging students with AI by saying, "These are the tools out there. We know you're going to use them anyway, even if we say don't use it, and so we want to help you, as an institution, to be able engage with these in a smart way." We are encouraging staff to think about both kinds of assessments not just lane two assessment, because then you won't know students know what they know, and not just lane one, because that would be very boring for students. We want to encourage those two approaches.

The CHAIR: Professor Bridgeman, do you have a contribution?

ADAM BRIDGEMAN: We're obviously both from the University of Sydney so it's a similar answer. I would add that our accrediting body, TEQSA, has asked us by June to have a plan and it will take us, I think, two or three years to get there because we have three-quarters of a million assignments, assessments and essays submitted every year. They're all very doable by ChatGPT so we need to sort that out. We do have a lot of sit-down exams, which was partly as a response to the outcomes of COVID, but we don't want to stay there because having just exams doesn't really prepare students for the future world.

The Hon. Dr SARAH KAINE: A lot of the talk about countering some of the negatives, particularly in the tertiary education space, seems to hinge on pedagogical innovation by academics. Academics are extremely busy people. With business pressures on them to get students through, what kinds of supports are we seeing, or not seeing, for academics to be skilled up? Because I see the list of potential examples you gave of how AI could be used in a good way. What is going on? Or what needs to be going on with academics to allow them to be able to be the first line of defence that we're talking about?

ADAM BRIDGEMAN: I think the first bit is actually exposing them to the possibilities of what the technology can already do. We've spent probably most of 2023 doing that, just showing them that their

assessments are completable by the technology. It is a two- or three-year process of professional development. We've got to get the tools into the hands of the educators—that's a money challenge—and we've got to give them time. I think that's the thing that our educators say they don't have, is time. We're lucky enough that we can have some grants to take people out of some of their roles—some of their teaching roles, often—and require them or give them a hand in working through their assessments, but it will take us two or three years.

DANNY LIU: I think it's a great question about true pedagogical innovation around AI. A lot of talk around AI for teachers in all sectors has been around efficiency gains and administration, but I think it's really important to think about how it can actually benefit pedagogy. What we're trying to do here, in advance of the vendors coming to say, "We're going to sell you this tool that can do X," to take Amber's point about all these tech companies saying, "This is how you should use AI," at Sydney University we're trying to experiment with different ways of using AI to put the control of AI into the hands of the teachers so that they can steer how these AIs can be used. One of the approaches is to get teachers to build what we call AI agents. That way they can build these agents which they can deploy to their students in their classrooms. They have full control over how these agents operate and, because they have full control, they can say, "I want this agent to be a role play agent for my students. I want this agent to give my students feedback before they submit." Because the teachers can control the narrative of how AI is used, we feel this helps them both upskill and also feel in control before the vendors can come in and say, "This is how you should use AI."

The Hon. CAMERON MURPHY: I have a follow-up in relation to the same area. Does AI present an opportunity, particularly in relation to assessment, to shift away from those traditional models of assessment, where it's a simple test or an exam into something new—something that perhaps provides a better outcome for teachers in terms of their workload but also prioritises students' progress and their wellbeing?

ADAM BRIDGEMAN: Absolutely. I think ideally our assessments in two or three years will be authentic and authentic to the new world, the changed world, so that it will be what we refer to as our lane two assessments. It's the assessment where AI, the copilot, is there naturally, just as it would be in the workplace. As I said, we have a huge number of essays. Academic essays are the bread and butter of many of our faculties. We have to move away from that because it's a skill that's changed. I wouldn't say it's irrelevant but the skill of researching information and collating information and critically analysing that information—the balance there in what the students needs to do has changed. It's fundamentally changed so we need to make sure that our assessments are doing that. We have a lot of multiple choice questions—pointless almost in that a plug-in can complete them for the students. So we have to move away from that and value the human things, the things that humans will carry on doing—the critical thinking—which what we say we do, but often our assessments don't reflect that.

AMBER FLOHM: On the assessment, the difference in the schooling sector vis-a-vis the higher ed is, of course, that we are dealing with much younger people, children and young adults. So the agency is quite different in a schooling setting. If a teacher uses a particular tool and instructs students to use it, they'll use it. Whereas when you're dealing with adults there is a much greater agency. There is, therefore, a much greater responsibility on schoolteachers to actually understand what AI is, how to use it effectively and ensure all of those privacy and data and equity measures are met before they launch in.

In relation to, for example, the HSC—to go to your assessment point—obviously, NESA, the standards authority, has to have much more robust policies and procedures to be able to address those assessment concerns that all sectors have had. For public school teachers, we can't talk across 2,200 schools as one group. Some of our communities have very high use of AI by the students in certain parts of New South Wales. Others have not only no use of AI; they actually don't have the digital technology or the connectivity. So, in the public school system, it's quite distinct. You have to address those variables to ensure we don't entrench disadvantage across the system. That is a significant concern for the teaching profession.

Ironically, the federation is engaged in a research project with the University of Sydney on exactly these things. I was telling my colleagues about that earlier, and that has already shown a few thousand teachers have engaged in that research around the national framework for schools, which was delivered in December last year for implementation at the start of term 1. Obviously, I'm thrilled to hear that it would take the university sector three years because, for the public school system, there will need to be similar supports for teachers before then. But what the initial research has shown is that teachers are particularly concerned about losing their agency as professionals and, basically, having machines try to replicate what we see as a very intellectual and human endeavour—that is, teaching. That is a significant concern.

There is another concern around the system's ability to provide teachers with adequate professional development and the time to do that, on top of already unsustainable workloads, and also, within the department, expertise in AI. I'd feel very confident, given my experience working across a number of universities in New

South Wales and researchers in AI, there is extraordinary expertise in a whole range of areas. That is not something that is replicated to date. That's what teachers are looking for: some system support, professional development, addressing workload and, of course, ensuring those matters of equity are addressed in the first instance.

The Hon. CAMERON MURPHY: Thank you. I've just got one last question. What are the risks for educational AI models that are developed by private enterprise? I think in your submission you talked about the dangers of edtech. But I'd like to just hear quickly from all the witnesses about what those risks are and what, if anything, the Government should be doing to regulate against those risks in the future.

AMBER FLOHM: Sure. From the Teachers Federation and the public school teachers' perspective, we have, up until recently—that is, the change of government—quite a devolved system of education, which actually meant that providers of AI could knock on the principal's door and sell their wares. We must change that. We have to have really robust frameworks that ensure that can no longer occur. The Government must have full accountability and responsibility for which programs. That will be challenging, as the—

The Hon. CAMERON MURPHY: So, in effect, almost anything could make its way into a school, whether it was conscious of protecting privacy or not.

AMBER FLOHM: Correct. The federation has long put that to the Department of Education, and they've have been most receptive—and government—to that, that they have a central responsibility to check apps, have a quality assurance process that leaves the profession feeling safe for themselves but actually for their students and parent communities as well, in terms of data privacy and also those equity measures. Unfortunately, we have seen some really unfortunate experiences overseas that have done exactly the sorts of things that we would hate to see in public schooling, for example, around facial recognition and predicators of scores in assessment, that have entrenched the sort of inequalities that we already see. We must deal with all of those things first, and we believe that is the Government's responsibility, to ensure those policy frameworks are there well and truly before we roll out any of those sorts of apps. I will say that we are currently engaged with Sydney University—the Teachers Federation—and UTS, looking at exactly those things: equity frameworks for public schools, to ensure that, should you be a private provider, you will have to adhere to those equity frameworks before you can put any product to the Government, we would argue.

DANNY LIU: Everything Amber said, I think, was perfect. I love those focuses on equity, security and beneficence. I think those are exactly the things that higher education needs to take as well forward.

ADAM BRIDGEMAN: I think one of the dangers at the moment is there's new technologies coming out and more and more specialised technologies coming out. So it's just explosion, and it's not going to level off. Those changes aren't going to level off for a while. We don't know when. There's a tool for every discipline. There's a tool for everything that one might do in a profession. All the professions will have their own ways of doing it. So just the breadth of what we're dealing here is difficult. We're here representing education, but we should acknowledge that people have already put research data into these and lost that research data, data that was protected commercial data or sensitive research data. It's really quite easy for people to make mistakes or not realise what they're doing. PhD students trying to get some help with rewriting a paragraph suddenly enter the medical data that they've collected into the system. Some protections around that—when things are offered for free on the web, they're not free.

The Hon. JACQUI MUNRO: Thank you so much for coming and for your submissions. Obviously, we've got this national framework for schools now—but relatively recent. How have individual teachers or schools been monitoring or reporting into your organisation about how they are using AI? Are you getting a snapshot of the use of that technology statewide?

AMBER FLOHM: You're talking about the national framework into schooling?

The Hon. JACQUI MUNRO: Yes.

AMBER FLOHM: That was only agreed in December last year. What we know from the research that we've partnered with Sydney around is that teachers are unaware of that framework. It is a very high-level framework. It's actually not practically usable in a classroom every day. If you're asking how have teachers responded to that framework, the answer would be they haven't, because that framework was a guidance, governance piece for the State, for example, of New South Wales. Translating that now into something that is usable for teachers and leaders and school systems is the work that is now before us.

The Hon. JACQUI MUNRO: It sounds like teachers have been mindful to an extent of the principles that are discussed in the framework when they adopt that technology in their own classrooms. Do you have an idea of just how much schools and teachers are using AI at the moment across New South Wales?

AMBER FLOHM: Certainly, I don't have figures on that. But what I would say for such a broad and large profession—I think you'll find there's a really varied approach. Some people would not be using it at all, would not be aware that they have AI in other systems, for example, in their lives, and there are others who are probably using it every day. The concern remains, however, that they don't know what they're using with their students. And that is really critical. They don't understand how they need to protect the students and the data et cetera, and so that's why having those guardrails and frameworks in place before they do that—teachers use multitude of resources to meet very diverse student needs. One of things we haven't discussed is the really positive impacts of AI. We certainly know that many of our teachers, for example, of students with disability are using great technology in AI to assist the teaching and learning. But those guardrails are critical to protect everyone.

The Hon. JACQUI MUNRO: Do you think it's a top-down approach? Does a principal have the responsibility of implementing the principles or is there a specific staff member in a school that would be responsible for technology education, in a sense?

AMBER FLOHM: I think that would send chills through the teaching profession, the thought that one teacher would be responsible for AI in a school or, for that matter, technology. No, we believe it's a system's responsibility to resource the sector. You have to understand, of course, that we are significantly underfunded comparatively to other sectors. We do not have the resources to buy an AI expert. It's just not a possibility in the public situation currently. It is actually the system, the Department of Education, and that will have to start with professional development for teachers to understand what it is and how to use it effectively, and they will need time to do that. I think I said in my opening that anything that adds to what I would describe as a pretty volatile situation, in terms of the workload of teachers currently, would be a disaster. That would actually be far more far-reaching than just AI, I'm afraid.

The Hon. JACQUI MUNRO: So it's about getting the balance right. There are obviously lots of time-saving measures that AI can embed into a system, being able to utilise that effectively so that teachers can spend more time teaching, effectively, rather than having to do things that maybe can be automated. But I guess what you're saying with that systems approach, if I'm right, is that every teacher goes through a professional development course. Is that right? Is that the goal?

AMBER FLOHM: Certainly we would want to see professional development rolled out by the Department of Education during staff development days, which are exactly for that purpose. That may give teachers an opportunity to start having discussions and actually learning from their colleagues who are already using it and are more experienced. That's how those things develop. I believe that is on the cards moving forward—but that's just the start. You actually have to have the digital access. That is a major issue for the Government to overcome. Some kids already have access to apps—and yes, you're right, they're not free. Other students do not have access or their parents do not have access. They don't have mobile phones. You have to ensure that the equity of engagement and distribution is met in the first instance, because otherwise the impact is on those who are already experiencing intersectional disadvantage. We will not have that as a profession.

The Hon. JACQUI MUNRO: I understand. To perhaps all of you, I guess there's a difference between using AI for specific curriculum-based outcomes and then, I understand, there are some mental health apps and programs, for example, that are being rolled out in some schools to try to track the welfare and wellbeing of students. That's not necessarily specifically associated with the curriculum but it's still a tool that can be utilised to, ideally, assist students in their learning and their development. Is there any comment about that side of AI and technology use that's sort of extracurricular?

ADAM BRIDGEMAN: We get a lot of vendors approaching us with such tools and we've had a look at some of them. It's an area where it has to be especially careful. Because they're often, as I think Amber said earlier on, trained with data from a different continent, one would have to be really careful. We have got guardrails about making sure that these things are tested. It's very difficult to test them thoroughly. If you're going to give it agency to interact with a student who's got a mental health issue, for example, you'd really want to be very sure that the answers it was giving were not going to be counterproductive. But there are a lot of people coming forward with tools. We have not adopted any, I think, yet, but we've certainly been looking at it. As a tool on the side of a person, I think it's a good thing. But alone with a student, I don't think we're quite ready to do that yet.

The CHAIR: Can I just ask a question on that, professor? You said that there's a use for every sort of purpose, really, across the spectrum. How much of it is being driven by vendors? Are any of those vendors in, say, your university, in-house in terms of other faculties in the university saying, "Hey, we've got an app for you"? Is it just the private sector, or is it actually the research element of the universities themselves that are putting things on the table?

ADAM BRIDGEMAN: It's both of those things. It's definitely people in universities getting excited. People at university have been looking at these things for 20 years. People have been publishing in this area for

20 years. There is an explosion of people using them, training their own agents, as well as vendors—vendors all the way from, obviously, Microsoft and Google, who have a lot of money behind them, to start-ups coming forward. One of the mental health ones that I was thinking about was actually, I suspect, four people in a bedroom in New York. It didn't look a very impressive outfit. There's the whole range of people out there. It's an explosion of technologies.

The CHAIR: How robust are your procurement guardrails? How have they been developed? Are they done on an organisational basis or can you—are there guidelines that you develop yourself, because of the niche area, or is it guided by the institution as a whole?

ADAM BRIDGEMAN: It's in development. We have, I would say, very robust procurement processes for large pieces of technology. AI, of course, with the additional dangers and the additional power, we are developing additional ways of making sure that we're testing what we can and doing background checks. For smaller technologies, and certainly for the free technologies that somebody might come across on their browser, it's much more difficult for us to control individual use. Danny, anything to add to that?

DANNY LIU: I think one of the troubles right now is that because a lot of AI companies are going out there directly to teachers, teachers are able to try these premium tools for free for a few months and then use it with students. Then it becomes very hard for the university to say no to them afterwards because the teachers get enamoured with them and the students do as well. In terms of your in-house vendor question, I think it's a really good one. One thing that we need to do as research institutions is we need to get ahead of these AI companies and say, "These are the pedagogically meaningful ways we know we can research and prove that AI can be used in a beneficial way for students and teachers". If we wait for the vendors to do this for us, they're going to be selling us all these things that we don't necessarily want that aren't going to be good for students but will look good for teachers, will look shiny and will be very expensive for us down the track.

Echoing what Amber was saying before, the main issues right now, I think we feel, in terms of AI tool access and vendors coming at us are around access and familiarity. It's the access to the more powerful AIs. A lot of vendors do give you free versions of AIs but the free versions are built on things like GPT-3.5, which is the 1½- to two-year-old version of the AIs. They're not anywhere near as capable as even the AIs that have been released just this last week, which are very expensive to use. One of the things that we need to make sure that we do internally is that we need to make sure that whatever tool we give our teachers and students access to is state of the art, as much as we can and can afford. Without the access to the state of the art, be it from vendors or be it from in-house, teachers can't experience the capabilities of this AI and they can't keep up to date with what it can do.

The Hon. Dr SARAH KAINE: Sorry, I've got a very quick definitional question. Forgive me for my ignorance, and I can only see it once in your submission, but I've heard you now twice this morning talk about agents. I'm not sure whether that refers to a chatbot. Could you just explain a bit more what you're referring to? What's captured in "agent"?

DANNY LIU: Sure. From our perspective, an agent is an AI chatbot which has been given a particular set of instructions and resources to act in a particular way. One early example of an agent that we worked with—a lecturer from occupational therapy—was an agent which is an AI with an AI engine behind it, and that AI engine was given the instructions to, ironically, act as a kindergarten teacher. You are very busy. You have a heavy workload. You have tons of kids in your class. Your role as an AI agent is to work with an occupational therapy student and basically be their client, converse with them and work with them. This AI agent was given the instructions and the information from inclusive practice and everything that the department gives out and to act in conjunction with the student.

The student would be sitting in class. They would be talking to this agent who was called Mrs S and say, "Hi, Mrs S, what can I do for you today about your student Jake." And this agent was able to be Mrs S and actually interact in that role-play with the student. The pedagogical purpose of doing that was to give the students access to apply their theoretical knowledge a year before they would hit an actual client in the class to make their learning more authentic. That's possible now because of these powerful AIs that we can tell to act in particular ways. When you tell an AI to act a particular way, that's what we call an agent.

The Hon. Dr SARAH KAINE: Acting as your agent.

DANNY LIU: Exactly.

The Hon. CHRIS RATH: We've probably focused a lot on the risks of AI, but I was wondering—maybe, for the University of Sydney—we've already mentioned it a little bit with the Teachers Federation, but maybe you could focus a bit more on some of the benefits that you've seen already or the potential benefits of AI that you see being rolled out over the next decade or so?

ADAM BRIDGEMAN: I think the benefits for our students are that they're going to be using these tools for the rest of their lives in their jobs and in their professions. I don't think this one's actually come quite to fruition yet but, in architecture, for example, in the undergraduate architecture degree they don't get to three-dimensional modelling of houses. There is no time in a three-year degree, apparently, to get to the state where you can actually build the house. You can do the modelling, you can do the materials, you can do the engineering and the design, but they don't actually get to see the 3D modelling. Now, with an AI tool, you can do the 3D modelling in the first week and then, backwards, do everything. So you've actually got what the building is going to look like or the house is going to look like, and then you can pick your materials and do all the engineering afterwards.

The advantage for the student is they get to the final product almost at the start, and then everything is flipped, so it's much more engaging for the students. There is much less of the tedious stuff. They still have to do all the tedious stuff. They still have to understand all the shadow diagrams and all the lighting and everything else, but they're actually working with a virtual model of a building straightaway. So it's that kind of thing. I think Danny's example in occupational therapy—they're getting the more authentic experience that much earlier in a safe and protected way.

DANNY LIU: I might give two other examples specifically around feedback. Feedback is often an issue, especially in higher education, with very large classes and teachers who lack time. The literature shows that feedback is one of the key elements of actually improving student learning, so things that we're trialling now at the university with these agents is actually to have teachers design their own agents to give feedback. This agent is basically a helper for the teacher to sit with a student, digitally, 24/7. The student, before they submit an assignment, can submit this assignment to their agent first, and the agent can go through the marking rubric. It can go through the assignment requirements and go through and say, "It looks like this part of your assignment needs a bit of work. Maybe think about these things." The agent can be told to not write on behalf of the student and instead give this constructive feedback. The benefit of AI in this instance is to actually give students that 24/7 support in a very personalised way that students are really lacking. But, again, emphasising Amber's point, not replacing teachers. Teachers are in control of the AI to do their bidding, but in very small, fixed ways that are controlled by the teacher.

The Hon. CHRIS RATH: How do you think we mitigate some of the copyright and legal issues? I know you mention it in your submission, but are you able to expand on that a little bit more as well?

ADAM BRIDGEMAN: I think—and I mentioned research data as well—it's making sure that we are working in closed environments and, if we going to work with a vendor, that we have a very tight agreement about what they are doing with our data and where they got their data from, so they're not scraped. It is for us to make sure that we are dealing with vendors in the right way and then making sure that our educators and our researchers are using the tools that we've—as we mentioned earlier, had all those processes. We went through all those procurement processes to make sure that the copyright and the data is protected, and then we are using those in a closed environment.

AMBER FLOHM: I guess, in terms of copyright, that's not such an issue for young children and young adults. But I can only, again, emphasise the importance of the system taking control of those quality assurance practices and those frameworks rather than leaving it to a school-by-school decision. That's where we believe it will not only be very bad for teachers' workload and engagement in AI but actually most disadvantaging our students if there's not a systems layer. I did really quickly want to go to the benefits of AI for teacher workload, because we certainly wouldn't see apps for student wellbeing as a good thing for children and young people. We believe school counsellors and access to medical professionals are what our young people need, but, of course, that's quite distinct from adults who have a lot more agency. But there is real potential for the workload of teachers in admin tasks of no pedagogical or curriculum value, undoubtedly. I laughed about the multiple choice. For example, if there were AI systems in schools that marked multiple-choice et cetera, that would be excellent. There is no pedagogical or curriculum value in that. It's just a time-consuming exercise for teachers so, obviously, we would embrace that.

But there are other areas too. High school teachers in particular, with a focus on HSC, spend copious hours providing written feedback across essays—not always in red pen, I'm here to tell the Committee, but hours and hours of work on those senior essays. For example, if that data could be harnessed in a way that could feed through a reporting system to parents, that's really good potential. They'd actually be reports that would be useful to parents and students—that actually were contained in that a lot of cognitive load for the teacher during that time. So there's definitely potential to reduce the administrative workload too but, again, much like the university sector where it is actually the teacher, the professional or the expert controlling the inputs for student benefits. But we would be very concerned about data breaches in terms of student wellbeing and apps. That would make not only teachers but parents very nervous, I'm sure.

The CHAIR: Thank you very much for that answer, Ms Flohm, and thank you all. Unfortunately, that ends our session this morning. Thank you very much for your thoughtful and comprehensive contributions this morning and also in your submissions. We very much appreciate it. It will inform our deliberations.

(The witnesses withdrew.)

Professor EDWARD SANTOW, Co-Director, Human Technology Institute, affirmed and examined

Ms SOPHIE FARTHING, Head of the Policy Lab, Human Technology Institute, affirmed and examined

Professor LYRIA BENNETT MOSES, Director, UNSW Allens Hub for Technology, Law and Innovation, affirmed and examined

Dr KAYLEEN MANWARING, Associate Professor, Faculty of Law and Justice, University of New South Wales, affirmed and examined

Dr AARON LANE, Senior Lecturer in Law, RMIT Blockchain Innovation Hub, sworn and examined

Distinguished Professor JASON POTTS, Co-director, RMIT Blockchain Innovation Hub, affirmed and examined

The CHAIR: Good morning, everyone. Welcome to this hearing into artificial intelligence. We very much appreciate you taking the time to provide submissions to the inquiry and come along and give evidence. Do any or all of you have any introductory remarks to make? We'll start with Professor Santow.

EDWARD SANTOW: New South Wales has been acknowledged as a leader in artificial intelligence. We should build on that success and the State can't rest on its laurels. Reform to New South Wales law and policy is urgent to protect the people of the State from the harms that can arise from AI. I'll make two brief points. First, New South Wales has been seen as a leader in digital government. Australia was ranked number five in the recent OECD Digital Government Index, an achievement that drew heavily on New South Wales's approach. That approach recognises that digital government rests both on strong technology but also on strong legal and policy guardrails.

Digital ID is a good example. That initiative requires excellent design, development and implementation of the technology. New South Wales has shown leadership in that area while also taking advice from experts in civil society to make the technology more trustworthy and to make sure it doesn't leave vulnerable and disadvantaged people in the State behind. The Government has committed to introducing strong legal guardrails to protect against misuse, overuse and errors in the digital ID system, and the time for that vital regulation is now.

Secondly, the terms of reference for this inquiry reflect that the State's response to AI rests on a number of policies, and especially the NSW AI Assurance Framework. I have a connection to that work through my institute but also as a member of the New South Wales Government AI Review Committee. I consider that the Government should be proud of the praise it's received for its leadership, especially regarding the AI Assurance Framework, but any policy like this is only as good as its implementation. More is needed to improve the incentives and capability for all New South Wales Government agencies to apply the framework to ensure conscientious application of its requirements.

The Government has shone a light on a number of excellent AI projects in the State—rightly so—but as the New South Wales Ombudsman has shown, there have also been problematic instances of AI and automation. Federally, we also have the tragic example of robodebt. Those problems aren't trivial. They can cause real-world harm, including to human rights, and they can undermine trust in government and the underlying technology, so New South Wales needs to modernise its laws for safe and responsible AI. The Federal Government is basing its own reform in this area on the principles of accountability, transparency and testing, influenced by leading reports such as the Australian Human Rights Commission's report on *Human Rights and Technology*, which I led with Sophie Farthing in my previous role as Human Rights Commissioner. These would be good foundations for New South Wales reform.

The CHAIR: Thank you. Dr Manwaring?

KAYLEEN MANWARING: No.

The CHAIR: Professor Moses?

LYRIA BENNETT MOSES: Thank you very much. We would agree that technology-informed law and policy is essential to take advantage of the benefits of artificial intelligence while also addressing the harms. However, the main point we made in our submission is that artificial intelligence, or AI, is an insufficiently well-defined concept or stable category to be a regulatory target in and of itself. You can see this in particular with what happened in Europe and the various drafts of its AI Act, various definitions each time constantly trying to play catch-up as the technology was moving underneath it.

The current definition in the AI Act, if you take away all the optional elements—"may have this" and so forth—hinges on the word "infer". AI is systems that infer. That is a very confusing definition, I think, if you're

trying to look at computer systems and work out which ones are and are not AI. There is certainly a very significant question as to whether large language models infer in the sense used in the definition. If you look at something like the Federal Government's robodebt example, that was often critiqued as being AI. Yet when the Federal Government launched its AI ethics principles, the argument was made that those ethics principles did not apply to robodebt because robodebt was not artificial intelligence—which is technically correct. In other words, if one is trying to understand the category that causes harm, a particular technologically minded definition as a way to define the problem does not work. Yes, the law does need to address harms and potential harms associated with this technology, but it can do so far more effectively by starting with the values that we're trying to protect and going from there.

At the same time this inquiry is happening, we also made a submission to the New South Wales inquiry into the reform of the Anti-Discrimination Act, which does not have in its terms of reference issues associated with machine learning as something that discrimination law needs to look at. We would argue that's a problem. Discrimination law absolutely needs to address discrimination, whether those decisions are human decisions, system decisions or, as is most often the case, some combination of both. We would argue—and we argued in our submission—a number of other areas of law that we believe are values informed. Starting with values but with a technology-informed approach to change in law and policy is the most fruitful one.

The CHAIR: Thank you, Professor. Dr Lane?

AARON LANE: Thank you, Chair. Just some very brief remarks and then I'll hand over to my colleague Professor Potts. We are unashamedly tech optimists and just have a real belief in the power of not just artificial intelligence but innovation more broadly and what that can do for all Australians. We're engaged in a lot of discussions right around the country, not just in artificial intelligence but in other emerging technologies. In terms of our submission, we thought, given the breadth of the terms of reference, that it would be helpful to draw on our experience in speaking to a lot of different groups, businesses and governments in Australia and, indeed, right around the world.

Four topics keep coming up when people are asking us questions and so that's how we've framed our submission, talking about, "Well, what about hallucinations? Well, what about bias? What about the impact on work? What about competition in these models?" That's the context to our submission, but I thought I might hand over the Professor Potts to give a bit more detail about the background of the research centre and why we're interested in these topics.

JASON POTTS: Thank you, Dr Lane. The broader context here is we're the Blockchain Innovation Hub, and this is an AI-research-centred inquiry. The connection here is, in the past decade or so, we've had this super cluster of compute innovation. It's distributed computing and blockchain; it's deep-learning computing and AI; it's spatial computing and VR; it's Oracles and cloud computing. When you put this all together, what we have is essentially a digital revolution is happening on our watch, in terms of just fundamental new technologies that aren't just one technology; it's a bunch of them all happening together. What this is doing is this is fundamentally disrupting the economy. That disruption is occurring in businesses, it's occurring in jobs, it's occurring in tasks, it's occurring in organisations, it's occurring in the public sector and in the private sector. It's a global economic disruption that is coming through this stack of technologies.

GenAI is right at the core of that, because that's the workhorse one that is doing all of the big computer load. That's the one that you put on the user interface end, it's the one that you put at the backend of gathering all the knowledge. GenAI has a particularly important role in this economic revolution that is occurring right now. What we've been studying is how these new technologies get into the economy, and we're particularly interested in the role of government facilitating that, ensuring that the regulatory environment is fit for purpose to enable the enormous benefits that can come from these technologies to work through, to ensuring that those disrupted or harmed by the disruption that comes are adequately compensated or enabled to adapt to these new technologies. But a fundamental view is this is steam, this is electricity, this is an epoch-shaping economic revolution that is coming. Our job is to adapt to it as effectively and quickly as possible in order to get the benefits from that.

The CHAIR: It occurs to me from the submission from the Allens Hub that there needs to be coordination across government. I wonder if government is fit for purpose as it is now to actually do this work, the way we've set up, say, our ministries. I've never heard the term "digital government" used before. In New South Wales we have a Minister for Innovation, Science and Technology, but this is such a huge change coming. Do we actually need a new ministry to deal with digital government that coordinates all of our responses, working out if we actually want some of these technologies implemented and then coordinating the reforms across all the Acts that need to be changed from human rights to crime to property law—all of it? Do we actually need a ministry in this area to do this enormous amount of work? I'd be interested in any or all of your responses to that. I will start with Professor Santow; you look interested.

EDWARD SANTOW: The NSW Department of Customer Service, with the Minister being Jihad Dib and the previous Minister Victor Dominello, I guess have taken responsibility for that or have been assigned responsibility for much of that. Some of that responsibility has kind of flowed naturally through that customer service dimension because a lot of the work on AI is about providing services for the people of New South Wales. There is a really important observation, I think, in your question, which is that development and use of AI spans beyond customer service, right? While I think it's really important they've taken that central role, good coordination across the central agencies but also the other line agencies, I think, and departments is really important, and so perhaps more could be done there. It's worth noting that at the Federal Government level, there's a taskforce, which generally is sort of time-bound, but there's also an interdepartmental committee whose remit really is about that good coordination, taking good practices but making sure that if something is a problem in department X, it may well be a problem in department Y and Z, and so addressing that as well. The short version of all of that is I think there are really good foundations here in New South Wales and they could be built on to improve that coordination.

The CHAIR: Just on that, though, we had the AI Assurance Framework and the review committee. Talking to Dr Oppermann, that may not necessarily be going. What I'm saying is the national framework and the framework we've had in New South Wales has been time-bound. But from the submissions we've received, this is going to be a process that's going to continue forever, and so we probably need processes and a government response that exists in perpetuity if some of the changes that are coming are going to be so profound and ongoing and iterative as well.

EDWARD SANTOW: I think certainly in the medium term. There are technologies that are introduced and there's a settling-in period, and we're certainly experiencing that right now with AI. Then you stop thinking about the technology as how you deliver this. That is just sort of so innate to how you do the function or deliver the service. But I agree with you; in the medium term, I think that that is true.

LYRIA BENNETT MOSES: I have two answers to this or different ways of thinking about it. One is it's worth looking at the work of ANU's Tech Policy Design Centre, which is mostly addressed to the Commonwealth Government but I think the same kinds of principles apply about how government can be structured, without creating a new ministry as such but to coordinate across the different parts of government for whom different parts of this problem are relevant. The other one is thinking about the role that other kinds of institutions can play. It has historically been a really important function of law reform commissions, for example, State and Commonwealth, to keep the law up to date as new challenges arise. So that is one kind of body that can look at it. The challenge there is that they're very focused on law and so they don't necessarily have the sort of expertise in the technology.

In Europe, they use a device called technology assessment, and they have technology assessment agencies that do similar things to what in Australia, similar questions to what in Australia, is dealt with by law reform commissions but they're technologists looking at it and not lawyers. That means they get a different part of the complexity understood but miss some of the more legal technicalities. Being able to put those things together, the technical expertise and the legal expertise, into ongoing processes to address the challenges, not just of artificial intelligence but indeed of technological change and its implications for law and policy as that comes up in different spheres—and blockchain would be another one, for example, but there are many more—would I think be really useful and creating a body that could do that and have that mission. I have done some work on that in my own writings. I'm happy to send a copy to the Committee if that would be useful.

The CHAIR: That would be greatly appreciated. Does anyone else have a contribution?

AARON LANE: I'm happy to take another different tack. I'll leave what particular frameworks are in place at the moment and barriers to others. What we can provide some insight on is into that knowledge diffusion, innovation diffusion mechanisms. That is something that's usually lacking. Government does top-down really well; it does bottom-up terribly. We've both studied public sector innovation, which sounds like an oxymoron but it's really about what are the mechanisms that could be put in place to share innovation? In your last panel you were speaking about education. It could be mechanisms such as prizes or those sorts of things for teachers that share their innovative applications of pieces of technology. It's mechanisms for understanding, how are people actually using stuff on the ground and how do we feed that in and share that knowledge across the wider public sector?

JASON POTTS: To elaborate on that same point, I think GenAI is a general purpose technology. It doesn't sit in any one industry; it sits across everything. But it is also very specific in the sense that its uses in one area are not its uses somewhere else. As my colleague was alluding to here, the real challenge to get the productivity benefits from it is to incentivise people who learnt something in one part to share that information with someone else. We have a situation where there are very weak incentives to do that. We have the sort of

intellectual property system as one way of doing that; that's not appropriate here because of just how specific and fast-moving everything is. We have universities which do that, which will do that but slowly.

What we need basically is some sort of cross-departmental learning forum mechanism—whatever that is—to incentivise ideas and insights, including problems that have arisen in one thing to be shared quickly across the public sector in other things. The reason to do that is adaptation. You are wanting to deal with the disruption and learn quickly and adapt to it. A ministry is a long-term solution. What we need is fast, short-term abilities to move information and knowledge and experience as quickly and effectively as possible to where it needs to get to. That's a hard problem to solve. That is one where in every individual silo—"It's not my problem. It's someone else's issue to deal with." So some kind of horizontal learning mechanisms are what is required.

KAYLEEN MANWARING: Just to add to issues around technology assessment and to add to some of the other stuff that the Blockchain professor has just said, you perhaps already have a lot of the tools you need to help—the ministry will have to be long term, but it will have some short-term projects. Some of this can be done even across disciplines and has been done. For example, there has been the Horizon Scanning Series done by the Australian Council of Learned Academies, where they received some Commonwealth funding to look at particular issues in emerging technologies. I, myself, was involved as a consultant in the Internet of Things Horizon Scanning Series. They produced a report, for example, that covered a whole lot of issues. A ministry could be used, then, to coordinate, essentially, adopting some of the things that happened—rapid response reports being done. In the sense, I agree that you have to have both the short term and the long term, but combining those two sorts of processes might be helpful to the government in that respect.

Ms ABIGAIL BOYD: Good morning to all of you. Coming off that question from the Chair, history has shown us that you can only rely on business to an extent to share its expertise when it comes to regulation. Some of this is giving me memories of the financial crisis, when we found that regulators were faced with such a complex scenario. There were people in business and people in finance who fully understood everything that was going on, but when it came to the regulators, they were behind and not able to keep up in order to regulate properly. I worry that we're in that situation here. If we were to pump up the public sector to be able to have more of that expertise itself so that we would be better regulators, we're faced with the problem—at the moment we're spending squillions on IT consultants because we don't have that capacity in-house. Have you seen this done anywhere else? Are there any good examples of where government has built up its own tech expertise, where it can then more adequately regulate? I will start with you, Professor Moses.

LYRIA BENNETT MOSES: I was going back to the example, in Europe, of technology assessment. In Parliament in the UK they have POST, which is the Parliamentary Office of Science and Technology, if I'm getting my acronym correct. That advises the Parliament on technical issues. They tend to hire people who've got PhDs in relevant areas of science and technology and who are interested in taking on a policy position and being involved in the question of policy around technologies. So there's absolutely precedent for it. If you go back into Australian history with technology assessments, you can find that we did have a body. I think it was ASTEC. I'll have to check the acronym. It was a similar thing but a number of decades ago. We haven't recreated that since. There are different models. There are some that are advisers to Parliament, so they are based in the parliamentary side of things. There are others that work more as units within government. But, either way, I do think that it's important to have one's own expertise, rather than relying on consultants, for obvious reasons. That certainly gives a more independent perspective to government than would come from industry consultation.

SOPHIE FARTHING: Can I add to that? The Human Technology Institute has been talking about this and bringing these two strands of questions together in terms of the regulatory ecosystem system and also what government is doing and what Ministers are doing. That concept of an ongoing, permanent advisory body that is independent and that will be a source of expertise, both to government and to the regulators, and the way that we are looking at the challenges posted by AI—there are multiple different strands of action that will have to happen. But one thing we are strongly advocating for is an advisory body, along the lines of what you're saying, to give support to regulators, who are all mobilising in different ways.

We are having a lot of different strands of regulatory action in New South Wales. I am referring to the NSW Ombudsman, who released their report on Friday. There are strands of action happening everywhere. There are strands of law reform that are being proposed at the Federal level, certainly. What we have been advocating for is bringing those together, strategically. Also, a really important part of that is an ongoing, permanent source of independent expertise. I think we've probably seen the first strand of that with the temporary advisory group that has been set up at the Federal level by the Department of Industry, Science and Resources. But we want that to be a permanent body. We think that will be beneficial, both at State level and certainly at Federal level as well.

KAYLEEN MANWARING: The other model that you do see, for example, is adding expertise directly to the regulators. For example, the United Kingdom Competition and Markets Authority has, in the last couple of

years, gone on a real recruitment drive to basically hire data scientists to help them keep up with the developments in different industries, particularly the technically heavy industries like financial services.

JASON POTTS: To build on Dr Manwaring's point about the horizon scanning idea, which I think is a magnificent suggestion, the universities are a magnificent resource for you on this. We've got huge numbers of experts across the areas, who are willing and able and want to work on the projects. A possible mechanism could be through the ARC, through linkages. For instance, one would be where you could put together short-term projects with a department and university partner to explore how things would be done.

The CHAIR: You said the ARC? That could be a model.

JASON POTTS: That could be a model. Obviously, that's not within your immediate remit. But the point is you've got a lot of resources already in-house in the universities. If a funding agreement can be arrived at to enable a lot of this expertise to be deployed, I think that would be, in many ways, superior to consulting models in terms of alignment of interests.

The Hon. STEPHEN LAWRENCE: I have a question arising out of the Allens Hub submission. Professor Moses, you talk, in part, about high-risk applications, and one of the examples you give is things involving the use of force. I was wondering if you could maybe talk to what a legislative prohibition would look like that was seeking to address that sort of high-risk application of AI.

LYRIA BENNETT MOSES: That particular example would probably be Commonwealth legislation, rather than New South Wales legislation. But my guess is the Government would have to think quite carefully about what exactly it wanted to prohibit in that particular context. The concern in a military setting, in my mind, is the idea of weapon systems ultimately making decisions for which there is no human accountability. And then something tragic and awful happens, and they say, "Well, I just relied on the system." There's no-one at whom fingers can be pointed out. The system itself might have multiple software developers involved and components and so forth. Everyone is pointing the finger at someone else, and there's no ultimate accountability. In that particular context, I think what's most crucial is less that it uses AI as such, according to some technology-specific definition of a particular set of technologies, but rather ensuring that, if you like, shoot decisions ultimately have a human who is accountable for them, and that that human has enough understanding of inputs from any systems on which they're relying that they are able to make an appropriate assessment.

The Hon. STEPHEN LAWRENCE: On that, and maybe for the panel more generally, the overall tone or substance of many of the submissions that we've received is that AI has lots of positives, and a lot of the risks relate to existing social issues, if I can put it that way. So it could accentuate social disadvantage or exclude certain groups et cetera. But I was wondering what you all would say about what some of the bright lines are, where certain things are just unacceptable, that we should be aware of?

LYRIA BENNETT MOSES: I think there is a difference between saying what are the bright lines and what are the bright lines that have to be drawn around a particular definition of a particular set of technologies. There are many lines that we need to draw. In discrimination law, for example, we draw a line and say you can't make a decision, say, about hiring somebody based on their sex, race et cetera. It seems to me that one has to craft that law slightly differently if what one's dealing with is not a human making a decision because of, say, race in that particular sense but rather a system that has drawn some kind of inference from data that leads it to, in its recommendation engine, point towards particular CVs that then correlate with an exclusion of particular racial groups, as an example. But if you were drawing the law to achieve that effect, you'd have to, I think, actually change what the law currently says if you wanted to achieve that effect consistently. So the red line is the same, right? The red line is around, ultimately, that hiring decisions cannot be on certain bases, irrespective of underlying data correlations or anything else. But achieving it is different.

So, I think, yes, I mean, I gave the military example as one where maybe there is a red line that is connected to a technology. You cannot leave a kill decision exclusively to technology, as such, independent of whether the word AI is appropriate or not. That is, if you like, technology specific. I gave another example in the submission about human reproductive cloning. We prohibit the use of particular technologies to create a human clone because we are concerned about the ethics of that. We don't ban the outcome, which is identical humans, because we don't kill one in every two identical twins. It is the technological means with which the ethical concern exists. I think there is an analogy to that in the military context where we might say an exclusively technology-driven kill decision is a problem. But I think in most areas where it is complicated, where decisions are made by human systems, the two interacting, we need to start from a different place than from the technology. We still need to draw a red line, but it's not necessarily one bounded by "Don't use the technology."

The CHAIR: Professor Santow, I would be interested in your response.

EDWARD SANTOW: I endorse everything Professor Bennett Moses just said—maybe two, quick additional observations. The foundation of New South Wales policy in this area—this is a really good example of bipartisanship—is international human rights law. That's explicitly stated, and so it should be, because I think what that does, and that's in line with the European Union's approach, it helps make very clear where those red lines arise. We don't have to rearticulate those things.

The second point is I think this is often really well understood by reference to specific areas where high-risk AI is being deployed. We did a big piece of work on the use of facial recognition technology, and one of the areas we focused on was the use by police. At the moment, you often have two camps. One camp says the technology should be utterly banned in all circumstances for police. The other camp says they should use it in any way they want, with no restrictions at all. What we said was a more sophisticated approach might be to identify where the risk might be highest, where the technology might be the least reliable and, therefore, where the human rights harms are most likely to arise, and really zero in on those things. And there would be some prohibitions there, but there would also be some uses where the risk can be addressed in advance to a level where the community is kept safe by a certain bounded use of the technology.

AARON LANE: When electricity was first invented, people used electric lights as the first application. That's a pretty simple and kind of boring application, but it was one of the first. It wasn't invented for things like other electrical appliances—refrigerators and TVs and gaming devices. The point I'm making is that, at the moment what we try to do is say, "Okay, what's a human task and what is AI going to replace it with?" And that's a good question. But the point is this: We're not going to know all of the tasks that AI in various contexts is going to do into the future in 10, 20, 50 years time, and so those bright lines are not going to be given; they actually have to be discovered. So part of this is an innovative process that—yes, there are some vigilance aspects to that as well, but there is also some experimentation that has to take place in order for us to say what those bright lines are going to be.

The Hon. CAMERON MURPHY: Just following on in the same vein, in effect what you're saying, I think, Professor Bennett Moses, in your submission is rather than focusing on the technology itself it's much more important, because of the fast pace, the different types of AI that are ill-defined, to be focusing on boosting that regulation over anti-discrimination law and other human rights law in order to make sure that those red lines aren't crossed and that the law is adaptable to dealing with the sorts of human rights situations that innovative design in AI might throw up. Is that right?

LYRIA BENNETT MOSES: Yes, and I think not just—I mean, human rights areas of law are certainly critical but, not only that. I actually think that there are a lot of areas of law that we need to review and look at that might have various different purposes.

The Hon. CAMERON MURPHY: Like consumer law, with the digital consumer manipulation and things like that?

KAYLEEN MANWARING: When talking about red lines, I agree that the point is to look at the conduct that is being enabled by these technologies. But one of the red lines that I think we have to look at, and this is important from a consumer protection perspective—both just normal products but things like insurance, which has a public benefit—and this will probably be relatively controversial but I think it's important, is we're talking about things we know, things we discover. We are limited in things we can know and discover because so much of this is controlled by the private sector, which has a cover of commercial in confidence, and that's something that government hasn't been able to pierce particularly effectively for big tech anywhere.

The Hon. CAMERON MURPHY: There are two things I want to ask out of that. Do we need some form of regulation that allows us to look behind the programming, so there is some level of transparency that lets us identify when those red lines have been crossed? Secondly, do we need to rethink the way we hold people accountable in the law? If you look at some of this, it may not just be the end user; it could be the developer at the other end of the equation who is at fault if it's an inherent discrimination or bias in a model. Do you have any views about that?

LYRIA BENNETT MOSES: I think transparency is a really complex question because I do think it's heavily context dependent. In other words, government, if it feels for its government systems—and I think for many government systems particular kinds of transparency are critical so that the public can have trust that the system will act in the a particular way. That can often be dealt with through procurement policies—it's definitely dealt with through the assurance framework—in forcing those questions to be asked. But I think it's very difficult to—I think a lot of the time when transparency gets talked about in the context of AI it's as if it's an unalloyed good and it's as if it's completely like all systems should be transparent. There are lots of kinds of transparency that are unnecessary or are not particularly useful for the kinds of concerns that people might have.

One example might be something like facial recognition technology. If, for example, government is proposing to use facial recognition technology for some purpose, there are things I would want to know. Does it, for example, perform equally well on people from different ethnic backgrounds, people who are different genders and so forth et cetera? But I don't necessarily get that information from looking at, say, the source code of the program that's running; I get that information from being able to test the system and being able to test it across a diverse set of faces, for example, in that context, and evaluate that system against particular performance metrics. And that is actually more important information for the public to have than just saying, "Well, show us the source code of the program."

So what is required and what kind of information people need to have about a particular system is going to depend on the context. One thing we mentioned in our submission was the idea of, if you like, government as a model user of AI systems and taking seriously the importance of particular different kinds of transparency when using different kinds of systems and insisting on that through mechanisms like procurement. Whether the idea that somehow all systems should be made transparent would be a useful one, I don't think so. I think it's far more context dependent. But, yes, there are circumstances, particularly in the context of government uses, where particular kinds of information about systems are critical to their appropriate use.

EDWARD SANTOW: Can I just add a point on that? I, again, agree with everything that Professor Bennett Moses said. In addition, there are two reforms that I think are really urgent in this space. Sophie Farthing and I did some deep work on this at the Human Rights Commission. The first is where you have a right to reasons because of an administrative decision. It would be completely undermining of that right if all you get, instead of reasons, is just the ones and zeros—the source code or whatever. You need to be able, if you're using AI to make a decision, to provide a meaningful plain English answer. We also know through generative AI that you can create a very plausible answer but it doesn't get to the true rationale for the answer so you also need to be able—it would be unusual, but in certain circumstances—to get a technical response to determine whether the plain English answer actually reflects the true basis of the decision.

That's a clarification of the existing law. It doesn't extend any new right to reasons in any way but it makes that right to reasons meaningful in the modern age. The second additional point, I think, comes from the ACCC's regulatory action that was taken against Trivago a couple of years ago. That was really crucial. It's a globally significant case. It's being talked about as much in Washington and London and Brussels as it is here in Australia. What they showed was, again, the same thing: You cannot maintain something as basic as the rule of law if you have a completely opaque decision-making system where you can never get to the basis of whether the law has been followed. Again, it's not talking about a significant extension of the law; it just means clarifying the law to acknowledge the fact that decisions are increasingly being made by AI and some of those decision-making systems can be a black box if they're are not designed well.

The CHAIR: Thank you. Professor Potts, do you have a contribution to make?

JASON POTTS: Yes. To follow up on this, just a word of caution about the scope of what can be regulated here because we've been dancing around this question of what these things are. They are not actually source code; they're models. A trained model is a just massively large matrix. You ask it to do things and it'll do things and they will be useful or not useful dependent upon the user. That's all the explanation you will ever get. There is no way to penetrate into why it did the thing. The builders of the model don't know why it did the thing. They will never know why it did the thing. They are beyond that complexity. The idea that we need explanations at all—we need to regulate something so that we can understand why it did something—will never happen here.

The Hon. Dr SARAH KAINE: What you are drawing a picture of is this disembodied model that hasn't been fed by anything. Surely there is not this thing devoid of any bias? There is something that creates that?

JASON POTTS: No, a foundation model, which is we are talking about here in AI, is a large, trained model that is trained on a dataset. That dataset is human. It's human cultural artefacts—language and so on. The bias that's in it is the same bias that's in the thing that it's trained upon. It's not uniquely biased. The reason AI is so powerful and useful is that that trained model is a general-purpose technology. You can use it for anything. You can inquire of it and it will do something. But this is also why it can escape and it can end up on devices. Lots of the open source models are now freely on people's devices. They're permissionless, in a sense. There's no firm or organisation that is granting permission to use them in a particular way, which is why the liability regimes are so complex and interesting here.

The Hon. ROBERT BORSAK: Back to you Professor Santow, are there any actual no-go areas, as far as you're concerned, for allowing AI into our lives?

EDWARD SANTOW: I think there are some situations where, if you can't provide for accountability and you can only offer a block box system then the black box system isn't ready to be used. A good example of that is in the aviation industry. For very good reasons companies—

The Hon. ROBERT BORSAK: What about in the democracy industry?

EDWARD SANTOW: Yes, absolutely. Facial recognition is another example. It is not enough to be able to go to court if you're a prosecutor and just say, "Well, the machine said that this person is the guilty party—the person who robbed the bank." You need to be able to produce much more careful evidence. Perhaps that links to what Distinguished Professor Potts was saying before because the way in which you interrogate these AI systems is not the way in which you would cross-examine a human witness. This is perhaps where he and I might diverge, but you can still get to what really amounts to an understanding of the basis of the decision. You use adversarial methods, and so on.

The Hon. ROBERT BORSAK: Should all AI code, for example, be open?

EDWARD SANTOW: Maybe, but I'm not sure that that actually gets you what you want. I think what we want is these systems, when they're making decisions that have a legal or similarly significant effect, to be accountable. You don't want to put too many barriers around that; you can innovate within that. Accountability, for those really significant high-stakes decisions, should be sine qua non. It should be an absolute requirement. Under our existing law it is an absolute requirement so we shouldn't be saying to the people of New South Wales, "We're going to wind that all back." What we saw with the Trivago case was the regulator there used a very different method to get to the bottom of the problem with that AI-informed decision-making system but it still worked. We still were able to get to the bottom of how a deep learning system was, in that case, misfiring. That's really important. It is absolutely possible. It's just a different way to how we would interrogate a human.

The Hon. ROBERT BORSAK: My question started thinking about democracy; that is, voting—what goes on in this place and all over Australia; the fundamental foundation of our democratic system. We have a bureaucracy that will be looking for AI-assisted voting systems at some stage, I'm sure. This place has seen computer-aided voting pretty much eliminated for the time being because it was a black box that no-one could understand and we weren't happy about it. How do you address this? We've had plenty of evidence over the years—certainly since I've been here for the last 13 or 14 years—that it all should be open-sourced but we just don't seem to ever get an example of that being put forward as a model that we should be using. We all want it, but can we trust it? My personal view is that there is no accountability in the systems that we've been offered in the past. They're not AI systems; they're just black box commercial systems that we're told to trust and then we find that they don't work. How do we interrogate? I'm listening to your evidence saying that it's just so complex and so deep—that's what Professor Potts is saying here—that no-one can actually understand them. Should the democratic process be ring fenced to say, "You cannot go in this space?"

EDWARD SANTOW: My advice would be slightly different. It would be to say that the thing about our liberal democracy is that there are principles that are absolutely watertight and they've stood the test of time.

The Hon. ROBERT BORSAK: But how do we hold those principles accountable in a system we can't understand?

EDWARD SANTOW: If you can never get to a basic understanding of how the technology works and it's something that is fundamental to our democracy, absolutely the technology is not ready to be used. People as wise in this area as Professor Stuart Russell have made that absolutely clear. He wrote a book that was one of the leading books on artificial intelligence but he has been one of the many people around the world who are advocates for the greater use of AI, who has said that unless you can understand how it works in a critical area, then the technology is not ready to be deployed.

LYRIA BENNETT MOSES: On the issue of electronic voting, talking to a fair few people in cybersecurity rather than artificial intelligence, there are a lot of concerns about all the systems that are used for electronic voting that are available at the moment. That includes the systems that are fully open-source and where you can access that source code in that case, although I take the point that in large language models it's a combination of the source code and a parameter model. My point is that it's not necessarily the question—just being able to look at the code and have someone look at it doesn't necessarily tell you, just by looking at it, whether or not it is secure and a safe means of voting. What all of these systems do is create systemic risk. At the moment there are errors in vote counting. Humans will occasionally put a vote on the wrong pile but those errors are not systemic. So, yes, every time you do the vote you'll get a slightly different count, but on average it's roughly right. What you get when you put it all into any kind of system is the risk of systemic error.

The Hon. ROBERT BORSAK: Or purposeful systemic error.

LYRIA BENNETT MOSES: Quite possibly—or someone from a non-friendly jurisdiction managing to suborn our systems. The point is that you absolutely have to interrogate that and test that from a security perspective before you ensure it's safe for democratic use. Now, there is a whole field of people—and I'm not one of them because I'm not a technologist—who do that, who test these systems and point out flaws in systems that are currently being used in countries around the world. But transparency can be part of it. You need to give the security researchers access to the system to do that testing, but I don't think that transparency alone is the solution to the security challenge we're fighting.

The CHAIR: Professor Potts, do you want to make a contribution?

JASON POTTS: Yes. In response to your question, the most likely use of AI in democracy is probably the uses that it has already been put to, which is campaign design and advertising—AI is a creative tool; you use it for coming up with slogans and things like that—and voters using it to answer the question, "Who should I vote for?" They'll ask their agent. We had a discussion earlier in the previous session about an agent. The use of an agent is to read all of the things. So I would say, "Agent, please go away and read all the things. Give me some people I should vote for and the reasons why." That kind of consumer use case for AI is a perfectly legitimate and effective use case. It enables a machine to do a thing that a human just wouldn't have time to do: to read all the literature and come up with arguments and use that. That's a perfectly legitimate use of AI in modern parliamentary democracy, without fundamentally changing the underlying rules and the process. It's just adding more information and computing into it.

The CHAIR: How do you respond to the issue of deepfakes, which have the opportunity to undermine the integrity of an image or a video that we see, and potentially undermine AI? How do you respond to that?

JASON POTTS: A genuine and real problem that doesn't just exist in that; it exists across a whole lot of areas. It's a technological problem that will be solved with just better technologies of detection.

The CHAIR: So there'll be an arms race. The only way to deal with that will be an arms race of detection and watermarking, or that type of thing. Professor Santow, you wanted to say something?

EDWARD SANTOW: I mean, maybe. We don't know. I don't know how you can give that assurance that in that technological arms race the good guys, as it were, will win. I just don't know. We're speculating. I think that the Federal Government has nominated that as one of the three areas of urgent reform. I think that's a good thing. But I'm not confident that the problem is going to be quickly solved.

The CHAIR: Mr Rath?

The Hon. CHRIS RATH: Thank you. I wanted to ask Allens Hub: You said in your submission we should consider the advantages of the Swiss approach to AI law reform over the EU model. I was wondering if you could expand on what you meant by that, or what is the Swiss approach? Why is it superior? Then, leading on to that just on the government regulation front, to RMIT, you've said thoughtful regulation, cautioning against potential regulatory lock-in—if you could expand on that and what the risks are if we overregulate AI, what impacts that would have.

LYRIA BENNETT MOSES: The Swiss approach is essentially the same kind of approach we're advocating for. Again, if it's useful for the Committee to have the paper that sets that out, I can provide that rather than rely on my memory of it at precise levels of detail. But, essentially, what they have done, or what they're proposing to do—because not all of it has yet been accomplished in terms of Parliament; it's more the approach that's being explored there—is to actually look at the different areas of law that need to be resolved. For example, if one takes liability as an issue, and there's some of the challenges in determining liability, then one might say, "Where are the question marks under current Swiss law, and in what areas is the current law okay?"—so the current law as applied to an AI system actually gives us the result we want, and then what areas do we actually still get uncertainty or it's not the desired result, and then limiting the reforms to those particular areas of law.

We're really saying, as opposed to the EU approach—you know, in Europe, there are plenty of laws that deal with particular things, with discrimination, with privacy, with other things, but it is like: If you use an AI system and that AI system is high risk according to these criteria, then you have to do these things. But if somehow it's not an AI system under our definition, then it doesn't matter if it's otherwise as high risk an activity: You don't have to worry about it. So where does the lock-in come from that? The lock-in comes from, if you like, the way that the technology is itself defined. You can go back and look at this problem of trying to define something technological, some phenomenon, and then regulate it based on the definition you have over the course of history of parliaments in multiple jurisdictions over time, and what you end up doing is you end up regulating the wrong thing.

An example that I liked to use from Commonwealth law a number of years ago was digital tapes. When digital tapes first came out and they made perfect copies, the music industry thought it had a digital tape problem. In fact, there was a whole section of copyright law that was added to deal with this digital tape problem. There was going to be a tax on digital tapes. The money collected from that tax was going to be redistributed to the music industry and so forth. The whole regime was set up. There was a High Court case about the particular tax power elements of it and so forth. But at the end of the day, they solved the digital tape problem. They didn't actually solve the copyright problem because they'd imagined a particular technological conception of what the problem was and did that. They were locked into that and when we had CDs and whatever was next after digital tapes, it wasn't under this new regime.

That's my concern, ultimately, with artificial intelligence, that we lock in a definition—and I don't think I've seen a good definition yet—that won't capture what comes next year any more than the definitions the EU came up with before the release of the later models of generative AI captured that. They're already having to have played catch-up over the time of drafting. You know, once they pass it, who knows what's going to come in that next year? What happens then is that industry plays games. If being AI is seen as a negative thing because you come under a different regulatory burden then, instead of innovating in general for good, they'll innovate around a definition.

The CHAIR: Dr Lane?

AARON LANE: Well, Committee, that's a hard act to follow.

The Hon. CHRIS RATH: Do you agree with all that? Do you agree with most of that?

AARON LANE: Yes, I would, Mr Rath. But if I might continue, by thoughtful regulation, in our submission, part of that is that it has to be functional rather than tech specific. If we're concerned about this, for instance, in administrative law, well, deal with that in administrative law. We don't need an AI Act. We don't need governments to build their own AI models. What we do need is to identify specific problems and address those specific functional problems that may or may not arise. A bad idea that is starting to take hold in other jurisdictions is this idea of licensing—that you need a licence in order to put out a large language model into market. That's a terrible idea for a few reasons. As Professor Moses alluded to, part of that problem is you get this regulatory capture where you then start either innovating around what the licence is, or what you have are existing players in the market lobbying to raise those regulatory barriers and prevent other innovative startups from coming into the market.

The CHAIR: And that's what we've seen in the States, isn't it?

AARON LANE: That's been the suggestion. But I think for competitive dynamics, you don't want to lock in a particular business model through the regulatory structures and prevent the next wave of innovation emerging. I think that there is a risk of lock-in in two respects: One is to have technology-specific laws that respond to a sense of urgency, where governments are called upon to do something, but that will have a limited shelf life and may cause other problems. But the second lock-in is if you lock in not a technology but a particular business model, that sort of deprives innovation as well.

The CHAIR: Thank you very much for that, Dr Lane. That concludes our time for questions and answers. I think we're left with more questions than we had before—my brain is full. But we very much appreciate the work that you all do, your taking the time to attend today, and the submissions you've made. If you've offered up any papers or suggested material that we should be considering, please forward those to the secretariat. They will be in contact with you in due course. Thank you very much for your attendance today. Safe travels.

(The witnesses withdrew.)

(Short adjournment)

Mr STEPHEN BLANKS, Treasurer and Past President, NSW Council for Civil Liberties, affirmed and examined

Ms LORRAINE FINLAY, Commissioner, Australian Human Rights Commission, before the Committee via videoconference, sworn and examined

The CHAIR: Welcome, everyone. We will begin proceedings. Do either or both of you have an introductory statement to make? We will start with you, Ms Finlay.

LORRAINE FINLAY: Thank you to the portfolio committee for the invitation to give evidence at this public hearing, and also for accommodating the provision of giving evidence remotely. It's greatly appreciated. I do seek the opportunity to make a brief opening statement simply to present the commission's overarching position with respect to the intersectionality between human rights and artificial intelligence, or AI. As Australia's national human rights institution, the role of the commission is to promote and protect human rights in Australia. As part of this work, the commission has a particular interest with respect to the complex interactions between human rights and technology, which includes assessing the human rights implications arising from the growing prevalence of AI in Australia. Our statement today builds upon the previous work of the commission to advocate for human rights-centred design and deployment of new and emerging technologies and emphasises the importance of responsible and ethical use of AI.

There is no question that artificial intelligence has been and will continue to be enormously beneficial, allowing for enhanced efficiency across multiple industries, with health care, education and finance being just some examples of industries where benefits are already beginning to be realised. At the same time, however, such innovation must come with adequate consideration of the significant risks and challenges that AI may present to human rights. One such challenge that we would like to highlight is the impact of AI on privacy, a cornerstone human right. As echoed in other submissions to this inquiry, the operation of AI may not only facilitate the invasion of privacy but potentially deepen these intrusions in new and concerning ways.

Another key concern that we've highlighted in previous work on the use of automated decision-making is the risk of entrenching unfairness and existing social disadvantages instead of resolving them. This potential for algorithmic bias may even lead to unlawful discrimination. The list of potential risks of AI continues to scale up with its increased use due to its widespread accessibility and user-friendly nature. This helps to facilitate potentially malicious applications of the technology, as evidenced by instances of deepfakes and the spread and increased impact of political misinformation. While there are clear benefits to be gained from the appropriate use of AI, there are also significant human rights risks and challenges that may cause real harm to individuals and communities without adequate safeguards and human oversight.

Our primary submission to this Committee is that adopting a human rights-centred approach should be prioritised when assessing the best strategy for using and regulating technology. This is especially critical when considering the application of AI in processes requiring critical human deliberation, such as within our criminal justice system and administrative decision-making. Ensuring that human rights are kept at the forefront of this discussion, placing humanity at the heart and remaining focused on human impacts maximises the benefits of AI while also ensuring that it's being used responsibly and ethically. Thank you again for inviting the Australian Human Rights Commission to give evidence. We welcome further questions from members of this Committee.

The CHAIR: Thank you very much, Ms Finlay. Mr Blanks, do you have an introductory statement to make?

STEPHEN BLANKS: Yes, just a brief statement, if I may. Thank you for the invitation to give evidence to this Committee. I agree with many of the others here that the community wants and should have the advantages and upside benefits of artificial intelligence applications, but the community also wants and should be protected from adverse consequences. That will require some regulatory and statutory changes. As to how to conceive of those, that is the difficult issue and one that you've heard lots of expert evidence about. I think it's clear that legislative responses can't be too prescriptive. They have to be principle based. We advocate for a risk-based approach.

I agree with Dr Finlay that a human rights-based approach is a good, sound approach. In New South Wales, the best way to secure a human rights approach is to have a human rights Act. We don't have comprehensive human rights protection in New South Wales, either at a legislative level or governing the way the public service agencies conduct themselves. The best way to achieve a human rights-based approach to AI is to introduce a human rights Act to protect human rights in New South Wales generally, and then it will follow that AI, along with everything else, will be implemented in accordance with an appropriate framework.

There will need to be created a statutory office to oversee AI developments, both as used by New South Wales government agencies and in the New South Wales community and the areas for which New South Wales

government has regulatory responsibility. Just like there is an Information Commissioner and a Privacy Commissioner, there will need to be a commissioner with powers to oversee, regulate and investigate issues arising from AI in New South Wales. Another fundamental principle that ought to be enshrined at a general level is a responsibility for any AI implementations to be fair and reasonable. That is a standard which the Commonwealth Government is currently looking at introducing in relation to privacy protection. As Dr Finlay correctly said, privacy is a significant issue with AI systems. Fair and reasonable standard ought to be legislated and applied.

To give an example of a recent AI issue that I came across—car manufacturers introducing in-vehicle camera systems to detect driver fatigue. The at least reported idea of these systems is to alert drivers when the system detects fatigue—the driver closes their eyes for too long or something of that sort. Those systems will have an AI aspect to them. Obvious questions that arise in relation to things like that: Will insurance companies demand access to the data that's generated, either in aggregate or in relation to particular individual drivers? Will law enforcement agencies—the police—be demanding access to that kind of information in relation to investigating accidents? Will mobile phone use detected by those systems be reportable?

Those are questions which are going to have to be dealt with, and they're going to have to be dealt with legislatively. I think what that indicates, along with many of the other areas you're looking at, is that the legislative response is going to have to be both at a general level and then at a particular sector level. There will have to be a legislative response in relation to the use of AI in education, transport, banking and finance, and they will each have different kinds of attributes to them. Those are just some general remarks I would make at the beginning.

The Hon. JACQUI MUNRO: Thank you so much for appearing today. This is a question for each of you in relation to the international situation that we find ourselves in. Dr Finlay, you mentioned misinformation as a very big problem to resolve. I understand that Google is rolling out a "pre-bunking" technology at the moment because they have also seen that this is a global trend and a globally identified problem. Are you aware of the role of pre-bunking—essentially, that you educate users of technology on what misinformation might look like so that they are pre-warned or have an understanding of how they might be duped? Do you think that type of technology can help with misinformation? Do you see it working in an Australian context?

LORRAINE FINLAY: Thank you so much for the question, and it does really highlight a number of important issues. The first is that looking at that international context really shows you the fact that, while we need an approach in New South Wales, it obviously has to be an approach that is made in full awareness of what is being done not only within Australia, but globally, because it's really helpful to have consistency amongst approaches in these things.

The second thing I'd note is the answer isn't a one-size-fits-all model. There are technologies that the technology companies are rolling out and developing that would be very helpful, but there also needs to be a response from government. There also needs to be a response that not only tackles misinformation once it's occurred but that works on primary prevention, effectively, by talking about education and giving individual citizens the tools they need to be able to ensure that they're aware of the information they're being presented too, and whether they can rely on it.

I think the other thing that's really important that's highlighted by misinformation are the human rights complexities that we need to give very careful thought to. Whilst, certainly, the commission recognises that misinformation is a serious problem, a growing problem and something that urgently needs to be addressed, we also need to ensure that we address it in a way that protects freedom of expression—which is another very important human right—and make sure that, in efforts to protect the community from misinformation, we don't start censoring different opinions and diluting the robust political discussion that we need to strengthen democracy in New South Wales and Australia.

The Hon. JACQUI MUNRO: I'm sure you're familiar with the recent scandal where images were being generated that were historically inaccurate and were causing a sense that what freedom of expression looked like from a technology point of view is different to what freedom of expression looks like from a human point of view.

LORRAINE FINLAY: Correct, and it's also important to recognise that, while the technology companies have a really important role to play in this, it shouldn't be left to Google to decide what's misinformation and disinformation. That's something that within Australia we need to think very carefully about—who gets to make these decisions, how do we know what decisions are being made, and making sure there's a transparency and accountability that attach to all of these things.

The Hon. JACQUI MUNRO: Do you think that there is a State role there, or is it primarily Federal?

LORRAINE FINLAY: I do think there is a State role there but, importantly, there needs to be collaboration and consistency between the States and Territories and the Federal Government, not only in relation

to this particular issue but across the board in terms of how we regulate technology. Of course, the technology doesn't recognise the borders that we have within Australia, and so we need to take a very pragmatic approach and make sure that everything we do doesn't just create additional regulatory burdens on business or additional complexities for citizens but is something that can work in a practical sense, so that we can get the benefits of this technology—which are enormous—but guard against some of those risks.

The Hon. JACQUI MUNRO: Do you have some specific examples of what that might look like from a State perspective? Are there pieces of legislation that should be amended to strike that balance?

LORRAINE FINLAY: I think one of the most important things that we have been advocating for is a regulatory gap analysis, so making sure that, in the move to better regulate AI or to ensure that we're providing the necessary guardrails around AI, we don't cause unnecessary duplication—that we look at the laws that are already there. Our preferred approach is not to regulate the technology, per se, but to regulate the use cases of the technology and to look at the harm that could be caused. There will be circumstances where there are particular risks or particular uses of AI that do need specific regulation, but our approach would be to start with that analysis, to make sure that existing laws are used where possible and then to strengthen and provide specific regulatory responses where there are those significant gaps.

The Hon. JACQUI MUNRO: Mr Blanks where do you see that interaction between that freedom of expression that was spoken about and privacy or misinformation?

STEPHEN BLANKS: That is a very real issue to consider. I agree with Dr Finlay that we shouldn't leave it just to Google to sort this out. Government needs to be active and have powers, but I wouldn't leave it to government alone either. I think it is necessary to empower individuals to take action where they've been harmed. Of course, we have very good examples in Australia of laws that give individuals remedies—for example, in the area of trade practices—for misleading and deceptive conduct. Yes, the ACCC has got extensive powers to take action against businesses that engage in misleading and deceptive conduct but, most importantly, individuals also have the ability to take action.

That is really what makes that system work, and so, in considering any regulatory environment for this space—I know that the Commonwealth Government is considering legislation aimed at misinformation online. But, in our view, what the Commonwealth Government is proposing falls short because they are proposing purely a regulatory regime empowering government agencies and not giving power to individuals. That is a fundamental defect, I think.

Ms ABIGAIL BOYD: Can I ask a question on the back of that?

The Hon. JACQUI MUNRO: Yes.

Ms ABIGAIL BOYD: Just on that, to ensure that human rights aren't being breached with something as technical and nebulous as artificial intelligence, the idea of empowering individuals is key. But is there a role or should there be a responsibility for government to educate individuals, courts, unions or whoever about artificial intelligence or tech more generally, so that people can actually see what's happening in order to then call it out and take the action?

The CHAIR: Do you want a response from both our witnesses?

Ms ABIGAIL BOYD: Yes.

The CHAIR: We'll start with Mr Blanks and then we'll go to Dr Finlay.

STEPHEN BLANKS: Yes, absolutely there's a role for government. I didn't mean to say that individuals should be the only ones able to seek remedies. There is a necessary role for government, both at an educational level and at a regulatory level in relation to the big corporates in this space. Each aspect plays its role in ensuring it. That's why you need principle-based regulation. You need a human rights framework that is enforceable, you need standards like fair and reasonable, which are flexible enough to encompass principle-based community standards, and then you will need specific sector-based things. I should have answered earlier, there will need to be reforms to particular New South Wales legislation. I note one of the other submissions, I think from Salinger Privacy, gave some very specific recommendations as to reforms to the Privacy Act and the Health Records and Information Privacy Act. Those, I think, are very sound recommendations but, yes, there are roles for everyone.

The CHAIR: Dr Finlay?

LORRAINE FINLAY: I should just say initially, I very much appreciate the promotion that the Committee has given me but I'm not actually a doctor. I just wanted to put that on the record.

The CHAIR: That's Mr Blank's fault.

STEPHEN BLANKS: Yes, I'm sorry.

LORRAINE FINLAY: No, not at all. I would agree with all of that but also make the particular point that I think that education really is key to this. It's education across the entire community, because one of the real risks with this technology is that some of the people that it can have the greatest benefit for are the people that find it the most difficult to access the technology. We need to ensure that we don't create a digital divide across New South Wales, or more broadly across Australia, whereby particularly vulnerable communities aren't able to access the technology or are more at risk from the use of the technology—in particular when we think about communities, for example, like CALD communities or when we think also about children, who are at particular risk in terms of some of the exposure to different artificial intelligence devices and perhaps don't have the same understanding in terms of the way that this information can be used and misused. It's really important to look at not just a one-size-fits-all approach to education but something that identifies education that can reach out to different communities within New South Wales and can have a layered approach in relation to children so that children, parents and teachers are all included in that process.

The Hon. CAMERON MURPHY: I have a follow-up question on that same area. Mr Blanks, you were saying that because of the risks and opportunities posed by AI, in effect it could be a catalyst for putting in place a human rights Act in New South Wales that deals with those principle matters, as you put it. Perhaps it's the opportunity to bring together discrimination legislation and other protections into one Act that's also extended to deal with AI. I wanted to hear more from you about that. Also, when you raise the issue of providing the ability for individuals, whether that's ordinary people or representative bodies like trade unions, do we need to extend that to things such as a tort of privacy, or even perhaps a tort of human rights, where people can go to the courts and take action in relation to those breaches that go beyond the technology but deal with those important human rights principles that we seek to protect?

STEPHEN BLANKS: Yes, thank you for that question. The NSW Council for Civil Liberties has long been an advocate for a human rights charter in New South Wales. There are many benefits to be had. Queensland, Victoria and the ACT have successfully implemented human rights legislation, and I think those States and Territories are better off for having done so. I think New South Wales would be better off for having human rights principles enshrined in a charter that governed the conduct of State agencies across the field and gave clear recognition to international standards in relation to basic rights such as freedom of speech, privacy and freedom of assembly—which, of course, current New South Wales laws are in contravention of—and a number of other important areas.

In relation to specifically extending that to the AI space, and what sorts of remedies should be available and who should have access to those remedies, when I said individuals should have access, I wasn't meaning to exclude organisations. Organisations such as trade unions or advocacy groups may well be appropriate to have the ability to invoke remedies. Whether the courts are the best place for the remedies to be sought, or whether there should be a specialist tribunal for these remedies, or whether the jurisdiction of NCAT can be expanded, that's something to be looked at.

It's necessary to ensure that the remedies are not beyond the reach of ordinary people who need them. We do have some problems with remedies in New South Wales in relation to privacy law. There's a ridiculous cap of $40,000 on the amount of compensation that can be ordered. That figure hasn't changed now for decades and is completely out of date. There's no reason why there should be a cap when the loss that can be suffered from injury to privacy can be much greater than that. Those are just some remarks.

The CHAIR: Mr Blanks, you were suggesting before that there should be a statutory agency overseeing this area, like the Information Commissioner or the Privacy Commissioner. Are there other jurisdictions that are moving to create and populate a role like that when it comes to artificial intelligence and its associated technologies?

STEPHEN BLANKS: I'd have to take that on notice, but certainly it's something which the Commonwealth Government is actively considering at the moment. There have been many recommendations to the Commonwealth Government that there needs to be a statutory agency dealing with these issues. Of course, one wouldn't want the New South Wales agency to double up or replicate the Commonwealth agency. But at least in respect of New South Wales government agencies, many of which will be using or are already using AI systems, then it'll have to be a State-created position to regulate or oversee those agencies.

The Hon. JACQUI MUNRO: Are there any international examples of regimes that you can think of that would be useful to consider in a New South Wales context? Taiwan has their anti-infiltration Act, for example, to try to stop the spread of misinformation from foreign entities. Obviously they have a specific threat, but are there other countries that we can be looking towards, in your view?

The CHAIR: Ms Finlay, would you like to respond?

LORRAINE FINLAY: Thank you, and then if I could also just make one brief comment about the previous question in terms of remedies for individuals. In terms of work in other countries, there are a significant number of countries that you could look to in terms of work that is ongoing, but I would say no one country that you can point to as being a perfect example of the response that is ideal, because this is an ongoing project and something that the entire world is grappling with in terms of the challenges that we face.

In terms of something like misinformation, even the initial question of how we define misinformation is something that, across the world, there isn't a single agreed opinion on or a settled definition of. It is something that is really being grappled with at a variety of different levels, not only at a country level but also at the United Nations level. There's significant work going on in terms of artificial intelligence and what the best responses may be, again, to ensure that we harness the benefits while guarding against those risks.

But if I could just make one comment, also, about the idea of a legislative response to effectively focus on remedies for individuals and providing individuals with the opportunity to protect their rights, that's obviously very important. But we would say that's only one aspect of a human-rights-centred approach and that, actually, what we need is a comprehensive approach that builds in human rights and thinking about human rights, not only at the end but from the development, the deployment and the use of this technology.

The reason that I say that is that putting the onus on individuals, for example, to protect their privacy rights is really difficult with this technology, when oftentimes they simply won't have the oversight of the way in which the technology is actually operating. Indeed, it's the prior impact. For example, with algorithmic bias, the training of the large learning models is where the difficulties actually can come in, in terms of setting up a model that then leads to unlawful discrimination or leads to unfair outcomes. For an individual to have to bear the responsibility for understanding how that works and how that impacts them, it's not a practical response. So there does need to be a lot of thought given to ensuring that, right throughout the life cycle of these technologies, we have the appropriate guardrails in place.

Ms ABIGAIL BOYD: Can I ask a question about the whole concept of consent and the way that it applies? For example, obviously we can decline cookies in our browser. But if I'm walking down the street and there's some use of facial recognition technology or something else, I'm not really actively consenting to that. That's often the response that we get. For instance, we found out in a transport inquiry that there was a form of facial recognition technology being employed on buses, and the response was, "Yes, but there are probably signs up telling people, so just don't get on the bus if you don't want to have your face scanned". I think that's not really an appropriate response. In the absence of individuals being able to actively make those decisions around consent in a lot of these cases, does that mean that the onus is more on government to regulate appropriately? We'll start with you, Mr Blanks.

STEPHEN BLANKS: Yes, I think that's one of the reasons why a fair and reasonable standard in relation to not just AI systems but privacy and the use of personal information generally is a necessary addition to the regime. Consent is necessary but not sufficient as a standard. Increasingly, as in the examples you've given, it's either absent in practical terms or is meaningless, because you have to tick the box because you want the service, and you don't have the right to negotiate the terms on which the service is provided. So there needs to be a fair and reasonable standard. That is being strongly advocated at the moment to the Commonwealth in relation to the Privacy Act reforms that are under consideration. I don't know where they're at in terms of taking that up. That might be an under-consideration matter, and it needs to be escalated to actually being implemented.

Ms ABIGAIL BOYD: Thank you. Ms Finlay? You're on mute again.

LORRAINE FINLAY: My apologies, you would think that I would have that sorted out after this long doing online interactions. You're exactly right that issues of consent do become quite complicated when you start to think through how people use this technology in their day-to-day lives. I'd make two points to follow on from what you said. The first is that it's really important that as the technology becomes more and more emmeshed with our day-to-day lives, and particularly around essential services, we look to provide realistic alternatives so that people actually can make the choice. If they don't want to engage with the technology, they actually do have the option to still be part of our community, day to day, and still engage with those essential services that they need.

The second thing I'd note is that it's not only a matter of a point-in-time engagement but also what can happen in the future, because we know that, with the data that's being held in increasing volumes, there are increasing risks about what happens to that data in the future. Who's it being held by? How's it being protected? Importantly, what other uses is it being put to? This technology is changing and evolving at a rapid rate. There are things that can be done now with data that only a few years ago we never would have thought about, and different conclusions that can be drawn about people in terms of the numerous data points that are now being

collected, which really can be quite intrusive in terms of allowing significant information to be held about a person and for that information to then be used in quite intrusive ways.

Ms ABIGAIL BOYD: One of the other uses by government that concerns me is the police use of facial recognition technology, and I asked during a recent estimates session whether the systems that were being used had been tested for bias. The answer I got back was that they had not, but no concerns had been raised, so it was kind of swept away. But it turns out that the technology that they're using is based on Cognitec, which I understand in the US has been identified as having some racial bias built into it. Does it concern you that we are using these sorts of technologies without having done the appropriate testing? Do you think they should stop using it until they have done that sort of testing? We'll start with you, Mr Blanks.

STEPHEN BLANKS: I think it's fascinating to suggest that police should stop doing something that they're currently doing. Whilst it's attractive to demand that, until these issues are addressed, I don't know how practical that might be. But certainly facial recognition technology, which is then used in conjunction with AI systems to formulate policy and formulate actions, is inherently problematic. Eliminating bias from the datasets may not actually be possible. It's certainly desirable, but I wonder whether it's not really useful to demand or to focus too much attention on eliminating bias from the datasets, rather than focusing on the process of decision-making based on the technologies and how the decision-making is potentially discriminatory, and direct the attention there. It's a difficult issue.

Ms ABIGAIL BOYD: Ms Finlay?

LORRAINE FINLAY: I think the short answer to your question is, yes, we are concerned. Certainly the Australian Human Rights Commission, going back to our 2021 technology final report, raised specific concerns there about the use of facial recognition technology in those contexts, given the significant impact that it has on individuals in terms of the harm that can be caused to them by the technology containing bias and leading to biased outcomes. In particular, we had a number of suggestions there about what could be done to address those issues, with the starting point for us actually calling for a moratorium on the use of facial recognition technology in those contexts, absent regulations being put in place to address the risks.

I say that in the full understanding of the important role that this technology can play in enhancing community safety and helping the police do the jobs that they're asked to do by the broader community. But we need to understand the risks that there are in this technology, and there is a lot of information out there dealing with the fact that there are concerns about the way this technology deals with bias, the accuracy of this technology and, as a result, the serious human rights implications that can flow from that.

STEPHEN BLANKS: I might just correct the answer that I gave earlier. Thinking about it, I think the NSW Council for Civil Liberties has called for a moratorium on the use of facial recognition technology until there is an appropriate regulatory system in place, and so that should be what I've just said.

The CHAIR: In the absence of any further questions, that concludes this session. We very much appreciate your submissions and you taking the time to come and give evidence today. It has greatly informed us, so thank you both for attending.

(The witnesses withdrew.)

Dr BENJAMIN KREMER, SC, Co-Chair, Media and Information Law and Technology Committee, NSW Bar Association, affirmed and examined

Mr BRETT McGRATH, President, Law Society of New South Wales, sworn and examined

Ms OLGA GANOPOLSKY, Chair, Privacy and Data Law Committee, Law Society of New South Wales, affirmed and examined

The CHAIR: Welcome to this hearing of the Portfolio Committee No. 1 inquiry into artificial intelligence. Do any of you have introductory statements to make?

BRETT McGRATH: Yes, I do, Chair.

The CHAIR: Excellent. We will start with you, Mr McGrath.

BRETT McGRATH: Thank you, Chair, and thank you, members. Thank you for the opportunity to give evidence on behalf of the Law Society of New South Wales at today's hearing. I'm Brett McGrath, the President of the Law Society. I am joined today by Olga Ganopolsky, who is the Chair of the Law Society's Privacy and Data Law Committee. Given the remarkable advancements in AI technology and its extraordinary transformative potential, it is incumbent upon us to carefully consider the legal and policy framework that will enable us to support the safe and responsible development and deployment of AI in New South Wales. The Law Society is closely monitoring the impact of AI on the legal sector and has instituted a dedicated AI taskforce to consider how the legal profession and the courts can respond to the profound challenges of AI as it continues to shape the delivery of legal services into the future.

We have long supported the development of AI laws that are flexible, scalable and principles-based to support innovation across Australian public and private sector organisations. The legal framework should, firstly, build upon and be adapted to existing laws and processes that Australian organisations have in place; secondly, be consistent with related legislation in areas such as privacy, cybersecurity, consumer protection and human rights law; and, thirdly, be cognisant of evolving regulations in other jurisdiction, including internationally, which might apply at various points in a data-driven service supply chain.

In our view, the fragmentary approach of AI law and policy in Australia represents a significant challenge for this inquiry, both in evaluating the current state of the law and in developing policy initiatives to promote safe and responsible AI in New South Wales. At the Commonwealth level, key law reform initiatives such as the ongoing review of the Privacy Act 1988, the 2023-2030 Australian Cyber Security Strategy and the safe and responsible AI consultation will have significant impacts on the state of AI regulation in New South Wales. Meanwhile, we note the key public sector institutions in New South Wales, including health and education, are already developing sector-specific responses to AI. We urge the inquiry to support, to the greatest extent possible, consistency in Australia's AI laws and draw upon the findings and proposals of the major AI reviews to date.

We also note that the recent trends in AI regulation internationally may provide useful insights for policymakers in New South Wales. In our submission, we drew attention to the European Union's Artificial Intelligence Act, which we note has since been approved by the council of EU Ministers and is currently awaiting European parliamentary approval. Since we made our submission, we also note that, in contrast to the more prescriptive regulations under the EU's Artificial Intelligence Act, the United Kingdom has, as of February this year, adopted a comparatively light-touch and pro-innovation approach to AI regulation, which provides a principles-based non-statutory framework for sector-specific regulations to be developed. This contrast in approach provides an interesting point of comparison for further consideration.

Finally, we welcome the inquiry's review of the New South Wales Government's policy response to AI under the New South Wales AI Strategy, assurance framework and mandatory ethical principles. We see significant merit in government acting as a role model in the adoption of ethical AI practices. Following our submission to this inquiry, the James Martin Institute for Public Policy published a report in December 2023 entitled *Leadership for Responsible AI: A Constructive Agenda for NSW*, which notes:

> … NSW … is well placed to take a leading role in addressing the impact of AI activity, for the benefit of our economy and local communities.

We suggest that consideration should be given to the proposal in the James Martin Institute report that the New South Wales Government's procurement framework could be utilised to help shape the market towards ethical and responsible products and socially beneficial outcomes. For example, this could be achieved by, firstly, integrating the AI Assurance Framework and the obligations it confers in respect of procurement into the NSW Government Procurement Policy Framework and, secondly, amending the New South Wales Supplier Code of Conduct to include elements of the AI Ethics Policy to clearly outline suppliers' responsibilities in respect of

responsible AI. The importance of ethical AI in the context of government is exemplified by the recommendations of the Royal Commission into the Robodebt Scheme, which notably proposed legislative reform to introduce a consistent legal framework for automation in the delivery of government services. Thank you, Chair, for the opportunity to give evidence today.

BENJAMIN KREMER: I have a brief statement. I also appreciate the opportunity to present and being asked to give evidence today. Our submissions are focused on two key areas relevant to our expertise. One is the potential use of legislation to protect the community, especially those who are vulnerable to the risks and harms likely to be associated with AI. The second is the potential use of AI in and associated with court processes themselves. Our overarching recommendation is that the Government should adopt a proactive approach to regulating AI, in particular to avoid misuse in what we've called "the frontier period", when regulation doesn't catch up with the speed of emerging technology.

We recommended a two-step process. The first step is the identification of any uses of AI that are so clearly undesirable that they ought to be prohibited by AI-specific legislation of general application. That's in contrast to, for example, legislating for particular sectors or industries where things could fall through the gap. The outcomes we have set out in the submission are, broadly speaking, the use of AI systems that are either manipulative, exploitative or perform social scoring that leads to differential treatment; facial or biometric recognition in decision-making or legal contexts; and the use of AI to replace or supplement judicial discretion.

Step two—and this is regardless of whether or not you accept our step one—is what we've called a regulatory gap analysis, where, in consultation with industry, sector and subject matter experts, you assess whether existing legislation is sufficient to address the foreseen, the hoped for or the feared impacts of AI and, where gaps are identified, to review and amend that legislation. We've given some case studies in our submission, including one we've carried out ourselves in respect of barristers' ethical obligations and AI or generative AI. Our approach, we think, is broadly consistent with the broad set of general regulations plus the sector-specific regulation approach proposed by the Federal Department of Industry, Science and Resources. We also see it as broadly in line with the position of the Law Council of Australia and the Australian Human Rights Commission. Finally, somewhat thematically separately, our final recommendation is to recommend the Department of Communities and Justice should review the use of AI in the court process and identify additional AI and related technology that could be implemented to enhance access to justice for court users with disabilities.

The CHAIR: I will start with a question to the Bar Association. You said that in your two-step process you are essentially creating a blacklist of uses that we should rule out. One of the guides is we should rule out manipulative processes, but it occurs to me that so much in our society relies on some degree of manipulation—if you look at advertising or, dare I say it, politics. How do we draw that line, in a legal sense, about what is manipulative and should be ruled out?

BENJAMIN KREMER: I think I'm cursed by my own attempt to summarise briefly the substance. It's recommendation 2 on page 13 of our submissions, if you did want to go to it in detail. There is probably no need to. The way we have drafted it is with an eye strongly to the EU's article 5. It's one of the prohibited practices that the EU has called out in its artificial intelligence law. The way we've worded it is:

> the sale or use of an AI system that deploys subliminal techniques beyond a person's consciousness, or purposefully manipulative or deceptive techniques, in order to materially distort their behaviour in a manner that causes or is likely to cause significant harm to them or another person …

with a carve-out then for approved therapeutic techniques. I think you've hit the nail on the head, which is it's very difficult to define with precision what it is prohibiting. But we note that very similar language is used, for example, in the Trade Practices Act—now the Australian Consumer Law—prohibiting misleading or deceptive conduct or conduct that is likely to mislead or deceive. If the idea or the concept of what is being prohibited is set out sufficiently clearly, it is something that will probably have to be developed on a case-by-case basis and, to paraphrase—I think it may have been Justice Blackman—you may not be able to define it but you know it when you see it.

The CHAIR: Interesting.

The Hon. Dr SARAH KAINE: I have a question for you, Dr Kremer, about your submission as well. I am interested in the second case study about using AI to replace or supplement judicial decision-making. Obviously, that has come out of concerns, it seems to me, about existing use or places where it has been used. Could you give a bit more information about the examples that might have informed that particular case and recommendations?

BENJAMIN KREMER: I think I may have to take the specific details on notice.

The Hon. Dr SARAH KAINE: Okay.

BENJAMIN KREMER: I think I can answer with that, so thank you for that. I think the general approach or the general thrust of that case study is there is an element of judging, whether it's decision-making, exercising a discretion or, for example, applying concepts such as reasonableness or proportionality, that—I think the view is—would be very hard if not impossible to replicate by automated means, by computerised means, whether it's called an AI or some other way of doing it. If you are going to have a situation where not just someone's finances, which could also be ruinous to them, but, potentially, their liberty is under challenge or is imperilled, there is a difference, I think, that we see between having a person who is accountable and who is trainable—there is an appeal process from. They must express reasons for their decision and expose their decision-making. There are complaints processes as well where one has that in contradistinction at the moment.

I think, to pick up Professor Bennett Moses's words from earlier this morning, when one has a black box where you have an algorithmic or large language model or something similar, where there is data and source code which is not understandable, certainly to the average person and perhaps even to those who are running the systems that become so complicated, it is not always understood how they work. If one is trading one for the other, we have hundreds of years of common-law process where a judicial decision-making technique is understood and has been accounted for, and moving from that accepted and tested system into a black box system, we see, has the potential to cause problems.

The Hon. Dr SARAH KAINE: Thank you for taking that on notice—I would be really interested. This hasn't come out of any Australian jurisdiction or judicial body? It's more the concern that this type of use of AI in the judicial process has been seen elsewhere?

BENJAMIN KREMER: I'm not aware of any Australian research or Australian examples of use or misuse. In paragraph 69 we do refer to some empirical work, I think done by ProPublica in the US, to do with sentencing, and that has the problem—I think it's also been adverted to this morning—where, if the dataset encodes biases or historic discrimination, it may not be possible to detect it in the dataset, and it may only be possible to detect it from the outcomes, in which case it's too late. I think that may, as far as I'm aware, be the only empirical data, but I will answer that.

The Hon. JACQUI MUNRO: To follow up on that quickly, the flip side of that is studies that show, for example, that decisions by judges or magistrates can be more harsh before lunchtime, when they're hungry, as opposed to afterwards, when they're feeling well rested and satiated. Is there a balance or a role for decision-making to be taken out of the kind of human foibles that might impact things like sentencing?

BENJAMIN KREMER: Again, a hard question to answer. I think Mr Blanks's answer earlier about use of technology that hasn't been tested or certified—it may also be worth noting that what we've currently proposed as being ring fenced areas or areas of high disapproval may change over time. For example, if someone comes up with a system that is validated and tested and proved, it could well be that this prohibition is amended or relaxed, and it can be allowed to be brought in. But I think one of the advantages of our proposal is you wouldn't have it de facto being trialled before any merits are tested. It would have to go the other way around.

The Hon. CAMERON MURPHY: I want to come back quickly to the prohibited AI practices. It seems to me that it's vitally important, Dr Kremer, that we have, if I can put it this way, a dynamic definition of those prohibited practices, rather than some static definition that identifies particular technology. Perhaps the exception to that might be something like kill orders from the military or some sort of police action, because the very essence of AI is that it is going to change and adapt, so we really aren't going to know what ought to be prohibited until we see it in practice, are we? You could get an AI model that starts off being quite ethical and useful and then, as it learns, it verges into something that we would want to regulate against because it's a prohibited practice, wouldn't we?

BENJAMIN KREMER: Yes—sort of a Skynet problem in *The Terminator*, where something goes rogue. I think we definitely agree with the idea of a dynamic definition, something that isn't so hidebound or restrictive that it's either left behind by technological change or even—again, I think Professor Bennett Moses said—innovated around. You sort of see the area of regulation, it's ring fenced, you avoid it, you see it in the rear-view mirror and it doesn't operate. I think the way we've tried to draft the recommended prohibited practices does try to be general in the sense of using language that is not very black letter and overly detailed. As I said before, the idea is that if you present either a regulator or court or whatever it is that is overseeing the Act, depending on the mechanisms set up for enforcement, with a situation where they see, for example, something that was intended to be a helpful aid is starting to take advantage of psychological tendencies, and it starts to fall within one of our prohibited areas—perhaps gambling type areas, where you move from advertising generally into an area where there is sensitivity—you would then be able to say this has fallen foul of the definition. The definition is broad enough to catch the change and it may also pick up new forms of misleading or subliminal messaging that distorts people's decision-making that isn't currently foreseen.

The Hon. CAMERON MURPHY: In a similar vein, is there anything you think we need to do in relation to liabilities or remedies? I'm thinking, for example, if you have a large language model or AI—say, a self-driving car—who do we hold responsible? Traditionally, we would hold the driver of the vehicle responsible. Do we need to look at law reform in the area of pulling that back to not just the end point but also the people who develop that AI and in some way regulate them or hold them accountable or responsible and then feed that into any remedy that might be available to someone who is wronged or harmed?

BENJAMIN KREMER: That's also a good question. I think there are a couple of parts to it. One of them, for example—to pick up on your example of self-driving cars—is that you obviously have current traffic legislation drafted around the driver. If, for example, New South Wales were to allow driverless taxis, like Waymo in San Francisco, that wouldn't fall within the current description or expression of the current laws. It may fall within the intended operation if there is a person in the passenger seat who is de facto controlling without a steering wheel and without hands on anything. That is an area that would be addressed by regulatory gap analysis where you would look at the traffic Act and you would say, "What needs to be amended to expand it from being the traditional concept of a human driver to some sort of automated driver?" I think that is within the sort of recommendation that we have made. I think it's recommendation three.

The broader point, though, about liability and redress is a very difficult one. We've drafted this as being a prohibition on the sale or use of systems. It would catch those at least who are selling it, even if they are not developers—certainly, developers who are also selling it. And use is obvious; it's the users of it. At the moment, the way it's drafted is to prevent the use of the systems. I assume they would be coupled with appropriate penalties or other sorts of legislation and legislative responses the same way you get in any regulation or licensing statute.

We haven't, I think, said anything about going further and moving into remedies to the extent that you're looking at third persons who are harmed, so not the State that is regulating or the entity that is regulating and the regulated entity but a third person who might be injured or have some other legitimate interest harmed. That would be something that would be worth looking into, although I'm not sure it would necessarily sit with a prohibition of practices. It could be, for example, that there's some kind of liability if someone is harmed from the use of a prohibited practice and it could also potentially even be strict liability, if you wanted to be very serious. Obviously, this is something about which reasonable minds may differ and there is no obvious answer. But one could look into some kind of right of action from someone harmed from someone else involving an illegal conduct.

The CHAIR: Mr McGrath, did you want to make a point?

BRETT McGRATH: Yes. To add to Dr Kremer's point there, firstly, for the member, you spoke about whether or not prohibited practices should perhaps be put in any reform. The Law Society has advocated that there is a flexible approach but a principles-based approach. The nature of regulation across the world is accelerating and, obviously, making our submissions outdated to a degree. California is a jurisdiction which I encourage the inquiry and Committee members to investigate. Obviously, that's the home of AI in that state. They are exploring options that may include things like kill switches. I understand that that's not formulated yet into any proposed legislation; they are still in the formative stages. But that jurisdiction, I imagine, would be the lead jurisdiction for AI tech companies in the Bay Area of San Francisco where these are coming out of. I encourage the New South Wales Government to look at that jurisdiction very closely.

On the point of automation and cars, I agree with Dr Kremer on that point. Again, we as the Law Society are encouraging the inquiry to look at the laws and regulations which apply at the moment and how AI and generative AI fits into the current framework. For example, for automated cars, if you've got a driver there, there's case law in—if I can take it on notice, the country, I believe, was Turkey. However, I don't want to give false evidence on that front. The driver in an automated car was found to be liable for the injury to the person. However, it arose whether or not he had a cause of action against the tech company that had the automation software within the car itself. There will be emerging areas which the courts are dealing with and which Australia as a common-law jurisdiction always looks to other jurisdictions of like—whether it be Canada, the United Kingdom or other areas; in India, for example, as well—which will inform the development of that. It may be that the jurisprudence is evolving, which falls within the current frameworks and statutory regime that we have.

The Hon. CAMERON MURPHY: It seems to me that the automated car is a relatively simple example. But if you have something like a private company which uses some machine learning program to automate a process that has inherent bias or something like that in it, you then start to get into a much more complicated process. Unpicking it, is there actually a way for government to try and regulate or provide remedies for people that apportion that blame or liability to compensate people for their loss who suffer through some form of discrimination or bias as an end user of one of those systems or somebody who is subjected to it?

BRETT McGRATH: To add to that point, I think the robodebt scheme is an example of, certainly, the liability the Government may incur if its departments and services are using AI and whether they haven't been put

through an adopted and formalised human rights framework or AI ethical use framework. Certainly, in the public sphere, the market may determine what occurs in respect of those remedies, but I take the point and I concur that the Government may face liability if it is utilising AI tools and hasn't done due diligence or applied a framework which is acceptable to a community standard.

The CHAIR: Just on what's acceptable to a community standard, we accept that governments make decisions about where to apportion resources and, on the basis of that, there are winners from that. People will say in the health sector they benefit because they receive funding for treatment and there's funding for the treatments that benefit them, but there's a sector of people who miss out. What if we hand those sort of decisions—like, we say, "Okay, there's an efficiency dividend in implementing AI in how we allocate resources in the health sector, but there's a group of people that miss out."

BRETT McGRATH: Chair, I again refer to the robodebt royal commission. An analogy can be drawn the same, that the Government allocates resources for social services and whether a veteran or a war widow was getting a certain pension or thereabouts. I just encourage the Committee to look at the outcomes of that royal commission because I think that is very instructive to that point for a government department that allocates resources on a day-to-day basis and makes tough decisions, which government has to; it's not a bottomless pit of money and resourcing. However, where does the culpability and liability lie for government when it's applying a certain standard when using AI technology?

Ms ABIGAIL BOYD: I want to pick up on facial recognition technology by police, which is mentioned in the Bar Association's submission in some detail, which is great. I've asked this question previously of some other witnesses as well. I asked about this in estimates in the weeks just gone, around the hundreds of leads for investigation that were generated using facial recognition technology. Apparently the systems we're using are based on some Cognitech technology which was part of the study in the US that found that faces of people of colour were 10 to 100 times more likely to be falsely identified than Caucasian faces. When I raised this in estimates and asked, "Are you concerned about the bias?", I was told there was no testing for that bias but also it wasn't a problem because it wasn't the only evidence used to charge someone. Does it concern you that people are being contacted on the basis of leads from facial recognition technology that could be biased even if that doesn't lead to actual claims? Can you talk about that? We'll start with you, Mr Kremer.

BENJAMIN KREMER: I think it's clear from our submission that it is something that the Bar Association is concerned about. That's why one of the prohibited practices is, based on article 5 to the EU Act, sale or use of an AI system that employs or facilitates facial recognition and other biometric technology in decision-making that has a legal or similarly significant effect for individuals or where there is a high risk to human rights, such as in policing and law enforcement. I think, at a conceptual level, it is something that is part of the Bar Association's submission.

I have to say that I'm not sufficiently across the underlying details, as I think I said earlier, to know how it's being used at the moment in New South Wales. I think, for example, if there is a practice that can be brought within that prohibited norm and the legislation is passed and passed in the way that we've currently drafted it, it probably would fall foul of that prohibition. Then, of course, one always has to have carve-outs, and the EU legislation deals with it differently. One of the carve-outs is usually national security. There's often also a carve-out for urgency or something similar. If you're searching for a kidnapped person and you know who they are and you want to use recognition technology at airports, for example, to see if they're about to be taken on a plane outside, I assume that is something that may well fall within that kind of an exception. Law enforcement is obviously the one that's probably going to be the most controversial.

Ms ABIGAIL BOYD: The argument from New South Wales police appears to be that it's okay because a human is involved before they get to the point of making an arrest. From a human rights perspective, or from concerns around over-policing, is that sufficient if we've got systems that have bias, but then there's an individual that gets involved before action is taken?

BENJAMIN KREMER: I think I can give you a general answer but beyond that I'd have to take it on notice. I think that the general observation would be that you'd have to look carefully at what purpose or what use the recognition software is being put to. Is it, for example, identifying someone for whom there already is an outstanding arrest warrant, in which case you're trying to locate a person but everything else that has already been done in the legal process has been done and it's just an extension of someone seeing them? Or is it being used to build up, for example, a record of a person's movements, and maybe not a specific person but everyone's movements, and then trying to work backwards from that in an investigation?

Even to my lay understanding, there's quite a big spectrum of what it could be used for and what could be being done with it. I think that would probably affect the answer to the question. Some of that may not be

objectionable but some of it may be. I think, to go beyond that into specifics, I'm going outside my expertise and I think that I'd have to take it on notice.

Ms ABIGAIL BOYD: I'll come to the Law Society. I think another example that I've have come across that I found interesting was that Transport were using smart CCTV, as they call it, to identify where there was trouble on train stations. You could get a situation where there was a group of kids actively up to trouble but then you could also get a group of kids who were just mucking around with each other and the question was what is the danger of police or transport officials going out to check? From a human rights perspective, or from the perspective of not wanting people to be unduly harassed or over-policed, is it sufficient if we have something in between an action being taken if that sort of bias is leading to people being contacted more than they might otherwise be? Have you got any reflections on this, Mr McGrath?

OLGA GANOPOLSKY: I'm happy to take that question and expand—as a very large body of work has gone into this thinking—and then take some of it on notice as well as to address some of the issues now. I'd say two things and invite the Committee to think about AI from a data perspective. What is it that you're regulating? I think, increasingly, the conversations and the depth of analysis that's being applied to this field is also asking us to think how we're regulating it. As you might have seen in other submissions or the discussion that's already been had, there are over 700 live discussions on how to regulate AI in distinction to automated decision-making. I think that while that sounds like it might be a word play, there are actually very substantial differences between what you would look at in order to make those decisions, and policy arguments would then be applied. I would certainly invite the conversation to move there.

When we look at that volume of activity, all of it is very thoughtful and all of it is from very well considered terms of reference but within the 700, the scope and variety leads you to ask those questions. Some of them go from an ethical perspective and some of them are highly industry driven—you've already had conversations about health, policing and administration of courts. So I think New South Wales, and certainly the opportunity that this Committee has, is to actually go a step further and think about how you're regulating this whole sphere and what is it that you're regulating. Some discipline needs to be applied to that. Of course, we're speaking here from a purely legal perspective, and we appreciate there are many other perspectives, but I think that line of inquiry is certainly emerging from the body of work that's been done overseas and in Australia.

The other thing that I would echo on the how we're regulating is that it's very interesting that every discussion that's been had is always very informative but always draws you to appreciate how different disciplines—not just different countries—choose to regulate that. Already there were some examples about law enforcement and consumer protection. We should be very clear in our discussions. Even the very considered and advanced EU AI Act is actually embedded in consumer protection even though we've applied it to so many other concepts. Whereas there's a whole field of, as many of you have already noticed, bias and risk that sits with the data protection regime and existing laws that are quite advanced in that field and, again, in all the industries that have a regulatory footprint. I am happy to provide further thoughts and supplementary material on that. I think it's important that we anchor the conversation in what is it that we're regulating. It's a much broader conversation than harms and a much broader conversation than the gap analysis to existing laws because those laws do speak to those issues already. The breadth of the discussion needs to have both the discipline and the transparency.

The Hon. JACQUI MUNRO: On those 700 conversations that are happening at the moment, whereabouts does that come from? How do you have that number? It would be great to get a list of those, if you have it.

OLGA GANOPOLSKY: I'm happy to dig that out and share it with you, and the references. Essentially it's a body of work conducted by a variety of regulators, both from ethical perspectives—so that includes data protection authorities in all the jurisdictions that have been considering this. We are happy to share with you that reference.

The Hon. JACQUI MUNRO: That would be very interesting.

BRETT McGRATH: Back to Ms Boyd's point around human rights in particular, the Law Society has long advocated a human rights framework to apply to New South Wales and also more broadly to Australia. Picking up on my colleague's point, it's not only what are you regulating but it's the starting point. When you look at jurisdictions such as the European Union or the United States, they have an individual rights-based framework which underpins their legal system. That frame of reference is restraint on government action, whereas Australia's system is parliamentary supremacy so it is incumbent upon Parliament to have to make these decisions and place its own limits on itself in that framework. That's why if there was a human rights Act that would provide that framework with which, universally, parliaments would apply. That's always been part of the argument that we've had but then further to that, how you address it now. In our submission, we've advocated that there needs to be transparency in the use of automated decisions. Whether that applies for investigations that are recurring use in

the justice system, what fundamentally underpins any use of AI is the transparency of its use and when it came into play, alongside the review process that may have been undertaken by a human.

The Hon. JACQUI MUNRO: What does transparency look like to you? Is it you've got a disclaimer basically saying that this decision was made with the aid of AI, or is it as detailed as having the whole algorithm associated with the decision-making that people can go and investigate? Where does the level really lie?

BRETT McGRATH: When you think about the—particularly if the State is using it, for example. The behemoth of the State against the individual is well known and that's why we have core principles underpinning the rule of law, such as the presumption of innocence and the right to silence and those things as well. Certainly in the transparency realm, when you have the individual against State actors, for example, whether they are the police or health making determinations, an ethical stance would be a proactive disclosure, rather than hidden perhaps on a website on page 39 that no-one visits. A proactive disclosure is transparency. We're advocating that. Certainly we're seeing that in guidelines and reports coming from different jurisdictions on how to—when solicitors and their obligations to the court to disclose to the court when they've used artificial intelligence or generative AI. There is an expectation certainly from the bench that that occurs and, certainly as part of that framework, a proactive disclosure.

The Hon. JACQUI MUNRO: Does that also apply to private organisations?

BRETT McGRATH: Yes.

The CHAIR: And potentially an opt in, so a customer of a legal firm might decide, given the option—do you want an augmented solicitor or a biological one?

BRETT McGRATH: We think they're good enough as they are.

The CHAIR: I know. We're about to find out. But is that the type of thing you're talking about—in terms of transparency, you give people the option to opt in or not?

BRETT McGRATH: Certainly, Chair, I can speak from the framework for solicitors in New South Wales, where we have ethical obligations, which are heavily regulated, in relation to our disclosure to clients, our obligations to the court and the community at large. Again, the guidance that we're putting out for solicitors and legal professionals in New South Wales is that those obligations have stayed true and held firm, whether it's the development from the letter to the facsimile to email to Google. Certainly AI falls within those frameworks. On the point of transparency and your ethical obligation in any disclosure and dealing with your client, it may incorporate, for example, in a cost disclosure that you may be using generative AI, for example. But certainly disclosure and transparency form a fundamental of any ethical framework moving forward.

OLGA GANOPOLSKY: Can I just very quickly add to that?

The CHAIR: Certainly.

OLGA GANOPOLSKY: Firstly—and again, this has already been mentioned. The Department of Industry, Science and Resources is currently having a live consultation on the framework, including the ethical framework, and has various questions about legal questions. The broader draft principle 6 deals expressly with transparency and explainability and that is the subject of some debate. If I could draw the Committee's attention to that, I think that that is the line of inquiry that then extrapolates what's just been said to a broader community. Essentially, it deals with explainability in a way that a human would understand what has happened under that system. It's not a simple disclaimer in the way that we as lawyers would understand; it's actually much more far reaching than that. On that note, I would want to invite the Committee to consider these issues taking back the comments that have already been made by the president about harmonisation.

It would be very, very counterproductive—certainly from a legal perspective—to have the standards embodied in different instruments that speak to different issues at different times. We really would reiterate the emphasis and the plea that this is an opportunity to harmonise and to create and, if there is an appetite to regulate, that that regulatory framework needs to be sufficiently harmonised or at least flexible enough to be principles based and technology neutral—that you're not going to have conflict with what are quite important standards. Again, echoing back on how much we're leaning on jurisprudence in other jurisdictions, with California, Europe, some of these issues are already defined. Some of these issues are existing jurisprudential norms. When we talk about automated decision-making, when we talk about human rights and the borrowing that we have from the EU, they come with a package deal, so to speak. We need to really consider that. I think this is a really valuable discussion to underscore the need for harmonisation. It's not a nice to have. In our view, that is an essential ingredient of a successful framework.

BENJAMIN KREMER: If I could just add a few more comments on what the president stated, the Bar Association has also last year issued ethical guidelines on the use of generative AI for barristers. You'll see it's one of the case studies that we've referred to. I can provide copies. I've even got hard copies, I hate to say it, here if you want them. But what we've tried to stress is that generative AI—and if there are later forms—is just a tool, and it's going to be a tool that all lawyers are expected to at least be aware of. It highlights that the barrister remains responsible for the delivery of the services and your duty is not just of competence and diligence and to make sure you don't mislead a court but also duties of care to your clients to make sure that you actually do the right thing and do the proper research and get the right answer if possible.

If you point out the failings of AI, if you point out what it is, so you demystify, you look inside the black box and you try to say that it's literally a statistics algorithm that draws on patterns in the text it was trained on and outputs them randomly, sort of like autocorrect in a phone—if you understand that it's not doing any thinking and it is just regurgitating things statistically from what it's been trained on, you are then aware, for example, of the problems with hallucinations or outright fabrications and also the fact that it may give misleading or incorrect answers, not through that but simply through leaving something out.

The Hon. JACQUI MUNRO: I'm wondering if there are AI systems being tested in Australian courts at all at the moment.

BENJAMIN KREMER: I'm not aware of any being tested in courts. I am aware that the large legal publishers are obviously working on legal-focused AIs, because one of the great advantages of a ChatGPT-style AI is you don't need to be an expert in the query syntax you're doing—you need to define something in a database and for a lot of users it's a lot easier to write a natural-language, plain-English-style query and generative AI turns out to be very good at that. The legal research and legal searching area is going to see a lot of competition probably fairly soon. That is a very useful use of AI, again, as a tool that a lawyer can use, a barrister can use, a solicitor can use, as long as it's not something that they're reliant on and don't check.

The CHAIR: We very much appreciate you taking the time today to give evidence. It is very useful. We again thank you for your comprehensive and very useful submissions and recommendations. Thank you very much for both attending and your submissions.

(The witnesses withdrew.)

(Luncheon adjournment)

The Hon. PAUL LAKATOS, SC, Commissioner, Independent Commission Against Corruption, affirmed and examined

Mr LEWIS RANGOTT, Executive Director, Corruption Prevention, Independent Commission Against Corruption, affirmed and examined

The CHAIR: Welcome back to some of my colleagues and welcome to ICAC. Do you have an opening statement or an introductory statement you would like to give?

PAUL LAKATOS: I do. If you'd bear with me, I'll try and keep it fairly concise. Thanks for inviting us to participate in this parliamentary inquiry. As with many innovations and perhaps more pointedly with artificial intelligence, the advances bring both challenges and associated problems. It has long been recognised in law enforcement areas that the law has had difficulties in keeping pace with technological development, particularly, and consequently with the ability to legislate against and detect unlawful and improper behaviour flowing from the use of such technologies. The increase in evidence comprising digital data is well captured in a quote from David Ormerod, Chair of Criminal Law at the University College of London, when he gave in evidence to the House of Lords' Science and Technology Committee inquiry into forensic science on 6 November 2018. Mr Ormerod said:

> … in a simple street robbery 20 years ago … the complainant, the defendant, an eyewitness and a handful of others might have been able to contribute some small scientific evidence. Now, the potential is for that evidence to be: that of those witnesses, coupled with their mobile phone evidence identifying where they were at particular times; the messages from the complainant's and defendant's mobile phones; the dashcam footage from public and private vehicles going by; the CCTV evidence; the armband/FitBit-wearing defendant who can demonstrate that his heart rate was not escalated at the time of the robbery, and so on.

Mr Ormerod completed this by saying:

> The risk is that the investigators and subsequently the trial then drown in the data.

The drowning in the data is, of course, one of the big issues for organisations such as ours and others. The commission recognises the need to understand artificial intelligence technology, to identify its benefits and its dangers, and to apply it usefully to our legislative functions, which may be summarised as investigating corrupt conduct; examining the laws and practices applicable to public authorities and public officials in order to facilitate the discovery of corrupt conduct; and to educate and advise public authorities, public officials and the community.

The commission embraces the challenge to apply artificial intelligence to its functions and encourages other public bodies and officials to do so in assisting them to better prevent and detect corruption. Although outside the ambit of our work, the commission has noted other submissions, which appear to us to be sound, that particularly large private organisations and, more relevantly, those that supply public services must also embrace the challenge. We wish to emphasise that the commission can only speak of the impacts of artificial intelligence upon the work we do in the corruption prevention and detection area as we are, in effect, procurers of that technology and consumers. That notwithstanding, we recognise that if the commission does not keep up with such advances, the quality of our work will be substantially degraded. The commission, therefore, can speak to the application and effect of artificial intelligence in two broad areas: firstly, the capacity of that technology to assist in the detection and analysis of digital evidence; and, secondly, the ease with which such technology will enable people to subvert and contravene the law and other proper processes.

As with other law enforcement agencies and anti-corruption bodies around the world, a significant proportion of the evidence we now acquire—and therefore need to understand and process—comes in digital form that is the product of computers, communication devices and other like devices. For example, in line with international experience, digital data comprises 90 per cent of the property or evidence received by the commission. That has varied from 18 terabytes in the year 2014 up to 62 to 48 terabytes in the periods 2018-2019 and 2022-2023. Just so there is some handle on this for the people not so acquainted with computer technology and its capacity, 1 terabyte equates to 83 million pages of data. For every terabyte of data, commission staff have performed between 80,000 to 90,000 searches of that data. The quantity and size of the data acquired and requiring scrutiny is simply outside the capacity of the staff of an organisation such as ours to be able to deal with efficiently.

In November 2023 the commission established an AI strategy team with the purpose of assessing the current state of adoption of artificial intelligence and the gaps within our system, formulating and implementing a strategy, and facilitating artificial intelligence awareness training. Currently, the commission does use a number of software platforms incorporating artificial intelligence in the course of its analysis and investigation functions. The immediate application of artificial intelligence in the work of the commission will focus upon the management of complaints, of which we receive between 2,000 and 3,000 per annum. This is a labour-intensive task which

requires triaging, summarising and analysis of complaints to determine whether the matter should be pursued and, if so, how.

The second area in which artificial intelligence is applied in the commission relates to the redaction of irrelevant information from documents which are to be released publicly—that is to say, in the course of a public inquiry. The second substantial area in which artificial intelligence impacts upon the work of the commission is in the potential for such technology to facilitate bad actors in engaging in corrupt conduct and subverting or avoiding their legal obligations. The commission has not yet dealt with any allegations where it was contended advanced technology was used to perpetrate serious corrupt conduct, but we believe, with the significant steps in innovation being already taken, that will occur in the near or not-so-distant future. Preventing such activity in areas related to corruption and investigating matters as well as attributing responsibility for such conduct to identifiable people, presents a further significant challenge to the effective work of the commission in the future.

To further the knowledge of commission staff and facilitate the application of artificial intelligence in our work, the commission, in conjunction with other corruption agencies, has become part of the National Anti-Corruption Investigation Network, which was established to share challenges and knowledge exchange in corruption investigations. There is a current proposal supportive of a joint agency study to unify research efforts in managing extensive digital data and to keep pace with worldwide technological advancements. In addition, the commission is in dialogue with an academic from a Sydney university to study the impacts of artificial intelligence on corruption prevention and investigations.

The commission is aware of and acknowledges the work of NSW Ombudsman, Cyber Security NSW, the Information and Privacy Commission, and experts from academia and others who are at the forefront of research and analysis of the development and application of artificial intelligence. We have read many of the submissions made to this parliamentary inquiry, each of which make substantial contributions to the discussion, and, in particular, we commend the analysis in the Australian Research Council Centre of Excellence for Automated Decision-Making and Society, which adverts to the many risks and benefits, more applicable to the areas in which the commission operates.

The CHAIR: I have a preliminary question. You said that you were working with the national anti-corruption network on a cohesive national response to AI and how you can use it and manage the risks. In your submission, you say there are significant risks, including hallucinations. We're not worried about those but the accuracy of data and analysis. How are ICAC and other agencies managing the use of AI and dealing with those potential risks?

PAUL LAKATOS: I think it's fair to say that we're still at the formative stages of all of this. At the moment each individual State body and the Commonwealth body are grappling with their own issues, and the object of this cooperation is that we all join together, hopefully identify similar problems and address them. The short answer is, how is not yet known. But those are questions that are very relevant and are on the minds of us all as we go forward.

LEWIS RANGOTT: We've just dipped the tip of our big toe in the water, Mr Chairman. We're not using it in any substantial way at all. We're adopting a pretty cautious approach. As the commissioner said, all anti-corruption agencies are dealing with this digital data problem. Standard, off-the-market e-discovery platforms all have a search function in them and, as these platforms develop, that search function will come with packaged-up AI capability. Whether that's marketing speak or it's real we're still testing out. But that's the logical place where some of our work will edge into.

The Hon. Dr SARAH KAINE: Thank you, Mr Rangott and Commissioner, for joining us today and for the submission. I have a question that goes to your introduction, Commissioner, and also goes to a question that I asked the ICAC committee on 11 December at a hearing here, when I asked how ICAC was looking at handling things like deepfakes. The answer I was given was that there was this establishment of the AI team and strategy, which you referenced in your introduction. Could you talk a bit about how far that team and that strategy have gone along and where that's headed? It seemed to have just started when I got that answer in December.

PAUL LAKATOS: I think it would be fair to say that we're still in the early stages and don't have the answers. I was aware of the question you asked the chief commissioner. To put it absolutely bluntly, we've formed the committee and the committee is considering the problems. That's as best as I can put it.

The Hon. Dr SARAH KAINE: So the basis of your submission was largely thinking hypothetically and starting that discussion process around the issues?

PAUL LAKATOS: Yes.

LEWIS RANGOTT: I might add that some of our sister agencies are a combination of a corruption agency and also a crime agency—in Queensland and Western Australia, for instance. So the deepfake problem is more likely to come to the attention of more traditional policing models. It is more likely those agencies will be on the front foot, and it's possibly more of an issue for our colleagues at the Law Enforcement Conduct Commission. That's one of the reasons for trying to work together across a number of different commissions. We haven't seen, in our own evidence collections, what people would think was a deepfake in terms of the Biden one, for instance, that has been in the press recently. That's not popping up just yet. At the heart of a lot of our cases, we see lots of fake documents, but it is more likely to be a fake invoice, fake CV or fake email. The ability of AI to produce better fakes of those types of things is coming, we think.

The CHAIR: Following on from that, in your submission it says that if harm becomes apparent, the black box nature of algorithms can increase plausible deniability in relation to the creation of these documents. Victims can become psychologically abstract and removed from the conduct, and responsibility can be deflected to the technology. Is that something we should be particularly worried about? It sounds like, in effect, the technology could create a firewall between a perpetrator and justice in a way.

LEWIS RANGOTT: We think it would make our job more difficult. Again, we haven't encountered this, but I picture the day when someone is sitting before the commissioner, they're shown a document and they say, "I didn't make that document. The computer did it." And we have to get to the bottom of that somehow.

PAUL LAKATOS: I think it's a very real problem, Mr Chair. It has been averted to, both overseas in the United Kingdom and also from various academics in this country, that when in the criminal law you have to prove that someone intended to do something and they, in fact, have used ChatGPT to write a document, does the intention go with the product? There is plausible deniability sometimes. Those programs, as I understand it—I've not used them so I'm not speaking from experience—can sometimes fill in extra bits which, if a person is not attentive to what is filled in, may turn out to be unwittingly false and not intentionally false. So plausible deniability, and pinning criminal responsibility and any other sort of responsibility on the person, might be difficult in those circumstances.

The Hon. CAMERON MURPHY: In your submission, I think it's on page 6, paragraph 2, you identified the risk that autonomous AI applications might, themselves, be involved in wrongdoing. What does that look like? Do you have an example of what that might be?

LEWIS RANGOTT: Not a live one in front of us. I can answer this way: Normally when someone comes before us and we are alleging they've engaged in corrupt conduct, they have a target in mind. They're only thinking one or two steps ahead. The idea of an orchestrated piece of software development that leads to corrupt conduct, that leads to a target—

The Hon. CAMERON MURPHY: A bad actor creating a program rather than a program itself going off into some sort of corruption behaviour. Is that what you're talking about?

LEWIS RANGOTT: Yes. If the program doesn't work—it doesn't do what it's supposed to do—that wouldn't be corruption per se, although we'd always want to look at that. The expression I've seen used is someone poisoning the well. They deliberately create a piece of AI that's deliberately biased or deliberately coughs up certain answers that author wants it to make. It's a bit of a hypothetical.

The Hon. CAMERON MURPHY: Would you see the role of the commission in the future expanding into looking at the developers of such programs as part of anti-corruption work that you do?

PAUL LAKATOS: I think that's probably necessary because, as Lewis was saying, both the program and the data that is input into the program, which might be biased or skewed, has a human source, clearly enough. Whilst the law is uncertain around all of this area, as I've indicated in the criminal law context, it does seem to us that often there might come a point in time when, if a program is doing bad things—to put it bluntly—and there needs to be an attribution of responsibility, it's the human actor who precedes it who may need to be held accountable in some form.

The Hon. Dr SARAH KAINE: Again, these might be suggestions at this stage, but we do have more government agencies coming in this afternoon. In your submission, you say that AI can assist agencies to identify corruption but is dependent on their capabilities. I was wondering what capabilities in particular you think the agencies would require to use AI to identify corruption?

LEWIS RANGOTT: The accepted wisdom is that modern technology, whether it's AI or not, is better than humans at spotting red flags, for instance. The red flag might be a suspicious invoice. If the computer program says, "Here's your suspicious invoice," you need to have the cognition to understand what the red flag really is

and what to do next. That seems to be the most basic level of understanding someone would need to deal with the outputs of the program.

The Hon. Dr SARAH KAINE: So the capabilities you're talking about are capabilities in using those kinds of systems or some kind of higher level, higher order capabilities?

LEWIS RANGOTT: More the former. I'm not really anticipating that a New South Wales government agency will be off creating its own fraud detection software. That will be on the market and you would buy it from the market. You would pay someone to point it at your data. Interpreting the results would be the main area.

The Hon. Dr SARAH KAINE: I'm also interested in procurement. We have a parallel inquiry in just about everything but one is about procurement at the moment. You talk about procurement and about an AI system used in Brazil to identify red flags in the public procurement process. I wondered if you could give any more information on that. You might need to take this on notice, but I can't see a reference for us to look up and get more detail about it. If you have any detail now, that would be great. If not, if you could provide it, that would be helpful.

LEWIS RANGOTT: I can provide the fine detail on notice but, as I understand that particular matter, it was relatively basic use of technology. It was comparing data in one pool over here and data in another pool over there and matching them and finding some red flags. I think in that particular case it was something that we see very commonly in our work. It was a public servant who was awarding contracts to their own company without disclosing the conflict. So that's rudimentary use of technology.

The CHAIR: How commonly do you see that? Sorry, that is outside the terms.

The Hon. Dr SARAH KAINE: That's for the other inquiry.

The CHAIR: That's for the other inquiry. I'll get the journalists down here.

LEWIS RANGOTT: Our world looks at corrupt conduct through a lens of matters reported to us as suspected of being corrupt. We see this quite often in our work, yes.

PAUL LAKATOS: The Brazilian model seems to have been developed by the World Bank. Obviously one of the big issues there was that their data pool was fairly extensive, so to set up such a system in the first place would be expensive and time-consuming. But that wielded much success in finding these various connections which otherwise might not have been found.

The Hon. ROBERT BORSAK: Mr Rangott, I might have missed this—I came in a bit late. You talk about it in your submission but how do you see AI helping the commission in its work? It's a major threat, of course, but how do you see it helping?

LEWIS RANGOTT: We think perhaps the main areas are, as the commissioner mentioned, the number of terabytes, which is more than a single one of our investigators could get their head around, so assisting in the review of that electronic evidence would be the main area. Similarly, we get in the neighbourhood of 2,500 heading up towards 4,000 complaints each year plus a number of other queries so, in some ways, automating aspects of that. A complainant might come to us with a 500-page complaint tagged with evidence and all sorts of bits of information. At the moment a human being goes through all of that and writes a report, so if we can speed that up that would be assistance to us. Even some of the basic things that the commissioner mentioned—when we put evidence on our public website, a human being, at the moment, manually redacts bank account numbers, phone numbers and email addresses and things like that. If a computer can do that for us, and we're satisfied with the quality, then it's probably going to do a better job than a human being. They are some examples.

The Hon. ROBERT BORSAK: Why do you necessarily think it would do better job than a human being?

LEWIS RANGOTT: When I say better, I might mean faster.

The Hon. ROBERT BORSAK: Thanks for clarifying that.

PAUL LAKATOS: And not only faster; it depends on the amount of data. If we can't get to it, then it doesn't get done. If a machine can, at least we've got to first base in that exercise. I understand now in the Federal Court and others they've got a technology-assisted review, which is a system that has enabled practitioners in those courts, the Federal Court and one other—I think Victoria—to use AI to discover and put forward discoverable documents. That now has some validity in the court process—a similar sort of exercise.

The Hon. ROBERT BORSAK: Obviously, you and the courts have got to use it carefully because you don't know what you don't know. AI is a learning technology but how do you know that it's learnt sufficient, or are you going to tutor it?

PAUL LAKATOS: Yes.

The CHAIR: The ICAC has investigative powers but do you have the regulatory powers, like enshrined in a statute, that you would need to unleash an AI bot to do searches? Rather than securing documents or a hard drive from someone, what if you're using a bot to go and search through files, scour the internet or those types of things? Does the ICAC actually have the powers, or does it foresee it will need new powers, to be able to unleash or utilise AI to do some of its investigative work?

PAUL LAKATOS: As I sit here and think about it, the answer is that we would still need powers of compulsion to get access to the information. There's been a recent amendment to the law enforcement digital access orders Act, which has allowed us and others to apply for warrants that would permit us to compel people to give up passwords and so on to their computers and get access to electronic data. So we need to get access to it. Once we've got lawful access to it, as I sit here it may be that we have sufficient powers to sift through that material. If it's open-source data—unfortunately a lot of the people we deal with don't necessarily put their material on open-source data. It would make it much more helpful, but I don't think that is probably such a big issue. Digital access orders are already here and, yes, we would be and already have been using AI-assisted investigative techniques to try to sift through the masses of material that we have.

The Hon. STEPHEN LAWRENCE: I've got a question arising from what is on page 8 of your submission about the risk of deference to AI. That's something that we've already heard a little bit about in the context of the robodebt issue and the post office scandal in the UK. I'm not sure whether you've heard about that or not?

PAUL LAKATOS: Yes.

The Hon. STEPHEN LAWRENCE: I am interested in your thoughts on how ICAC will properly and fairly deploy these sorts of AI applications that present a risk of undue deference in circumstances where courts, which have all the protections of the rules of evidence and disclosure and so forth, seemed to have failed, at least in the UK. In the context of an ICAC-type body where people are being examined in public in the commission but not with disclosure of material or rules of evidence, how are we going to protect people from undue deference to AI?

PAUL LAKATOS: I answer that question in the abstract because we've not got to it, obviously. The review processes that the commission undergoes, I've found out in the last year-and-a-half or two, are very extensive internally. Almost every decision—from some of the most small decisions, one might think, to the more serious ones such as public inquiries and who we investigate and so on—is run past all the senior people, including the commissioners and all the interested investigators who are investigating these things, as well as the lawyers. All I can say is that things will slip through undoubtedly in the circumstance you've put. But I really do think that in terms of how much human manpower we put into making sure that we don't allow any mistakes, including AI mistakes, to creep in, we are probably doing as much as we can do. That may prove inadequate, as it turns out. As everybody said, we're all on a learning curve here but it's a very real issue, I agree.

The Hon. STEPHEN LAWRENCE: Do you think there might be a role for legislative provisions that somehow speak to this risk of undue deference? It seems to me that in some circumstances—a criminal case, for example—it might be, in effect, a failure of disclosure for undue deference to occur, either at the investigation or prosecution phase, or in the course phase. I'm wondering how you could mould a provision that ensures that's not occurring? Early witnesses have talked about certain red lines or unacceptable risks—for example, in the application of the use of force by the State and so forth. Is there much that is more important than our judicial process and like processes?

PAUL LAKATOS: The answer is no. I don't think there is too much more important, but then I would say that, coming from where I've come from. The only thing I can think of—and, again, this is all developing—is that if you have to have a human putting his or her signature to a disclosure document who is answerable for proper disclosure and that doesn't happen, then at least we can attribute responsibility. It may not undo a preceding miscarriage of justice. I accept that, but I'm not quite sure how you actually do it otherwise. I haven't thought about it long enough, so I can't give a definitive answer.

LEWIS RANGOTT: I think also when we eventually put this slightly more advanced technology into operation, our investigators—our users of this technology—will have to be trained and understand that although they might not memorise the algorithm that is spitting out the results they need to understand exactly what the limitations of the technology are. They don't just press the button and it spits outs the culprit and the evidence. They'll need to intellectually engage with the technology that they're using and be trained.

The Hon. CAMERON MURPHY: On that note, have you thought about whether, as part of the use of AI in your investigative processes, it should be coupled with a degree of transparency where you explain to the public and to people appearing before ICAC exactly what you're deploying and how it's being used?

PAUL LAKATOS: There is benefit in that approach, subject to this rider. The courts have recognised elsewhere that if you disclose too much about various methodologies that you use, they can be subverted. It has happened in the United Kingdom and elsewhere where people have tried to extract deleted items from a BlackBerry phone. Ultimately the courts in the United Kingdom held that the gentleman that did it was not compelled to say how, because of commercial in confidence, which is the one that comes up often, but also because of the fact that they were fearful that those who misuse the technology can subvert these ways of discovering evidence. The answer is, in principle, a good one. There might be difficulties in application.

The CHAIR: In your submission you say that there is an increasing trend for data breaches. I think 20 per cent of the data breaches in 2023 that you are aware of involved social engineering schemes or impersonation. I'm aware of a major quite high-profile one recently—a Zoom meeting where a person was tricked into sending a large amount of money in a Zoom meeting online or something like that with fake personalities. You are saying that public agencies are already at risk and suffering from this. What do you mean by social engineering schemes and how does the Government protect its agencies against those?

PAUL LAKATOS: The social engineering schemes, as I understand it, are schemes which take the personal characteristics of a user using a computer and then wraps up a request or a demand in personal information, hence making it look like it's a genuine request by a genuine body. How you stop it is a question on notice. I think we are all grappling with that. I don't think we have encountered a practical application in ICAC. Again, it's not likely to affect our work itself directly. It may affect what the people we are investigating do and how that's done.

LEWIS RANGOTT: There are a couple of reasons to be worried about this. Because of the transparency of government operations, including this proceeding here, it's not that difficult to go and find out quite a lot of information about senior public servants and about politicians. All the words you have uttered in the Parliament are publicly available, so you could be targeted and we could be targeted in that way because of that volume of public information. The reason to be slightly less concerned is that, for any properly run government agency, it's hard to get the money out the door. You need to have a purchase order, get it signed off, talk to finance and get the money out the door. In some ways, small- and medium-sized businesses are the target there because, if you can fool the owner, they can authorise any amount of money going out of their bank account. I suspect the forces are targeting those organisations with the weakest banking and finance controls. That's not generally the public sector.

The CHAIR: We very much appreciate your submission and the work that ICAC does. We appreciate you taking the time to come here today and give evidence and answer questions. It has been very helpful. There may be some supplementary questions, which we'll get to you via the secretariat in due course. Thank you both for coming today.

PAUL LAKATOS: We are happy to assist where we can. Of course, it's a bit of the blind leading the blind at the moment for us, but we will do our best.

(The witnesses withdrew.)

(Short adjournment)

Mr PAUL MILLER, PSM, NSW Ombudsman, NSW Ombudsman's Office, affirmed and examined

Mr CHRIS CLAYTON, Chief Operating Officer, NSW Ombudsman's Office, affirmed and examined

Ms RACHEL McCALLUM, Chief Executive Officer and Information Commissioner, Information and Privacy Commission, affirmed and examined

Ms SONIA MINUTILLO, Acting Privacy Commissioner, Information and Privacy Commission, sworn and examined

The CHAIR: Good afternoon, everyone. Welcome to this hearing of the Portfolio Committee No. 1 inquiry into artificial intelligence. We thank you for your attendance today, for your service and for the submissions your agencies have provided. It is very much appreciated. Do each of the agencies have an introductory statement to make? I will start with you, Mr Miller.

PAUL MILLER: Thank you, Mr Chair. I would like to begin by acknowledging that we are on Gadigal lands and paying respect to Elders past and present, as well as to the children of today, who are the Elders of the future. As you know, we have provided a written submission to the Committee, so I'll keep the opening statement brief. I'll just outline some context about the NSW Ombudsman's work and particular focus in relation to AI. The particular jurisdiction of an Ombudsman is maladministration, which includes but is broader than unlawful conduct by government agencies. We handle complaints about it, we can investigate it and make findings about it when it has happened, and we make recommendations and try to provide guidance so that it doesn't happen again in the future.

In that context, in recent years we came to focus quite deeply on AI or automated decision-making—which can be one application of AI—after observing that there was a lot of discussion, debate and effort in government academia at all sorts of conferences and round tables and events about the future potential of AI and the kinds of AI capabilities that might be just around the corner, and exploring the potential harms and benefits they might bring and the kind of regulation that might be needed to control them. However, what we also observed was that these discussions did not tend to talk much, if at all, about the extent to which agencies were already using automated decision-making technologies. That was something that we ourselves only really began to fully appreciate when we started looking closely at a particular debt recovery system being used by Revenue NSW.

Our focus really has been on the here and now. What systems are agencies currently using or developing and how might that use involve or, better yet, avoid maladministration? In relation to that, there were two things that particularly concerned us. First was the general lack of visibility of what ADM was being used in government and what it was being used for. Although it was well known anecdotally that agencies were increasingly adopting ADM systems, there was not a lot of public information beyond a handful of well-known examples, which, in some cases, only came to public attention because something had gone wrong. We noted that agencies had no obligation to proactively publish details of ADM systems or inform members of the public when decisions were made with the support of them.

The second concern was what seemed to us to be a broad lack of appreciation among government agencies and officials about the legal frameworks that already existed and how they controlled the use of ADM. It was almost as if there was a view that, because the technological tools were new or because they were not expressly prohibited or regulated, their use was unregulated unless and until new laws were made specifically to deal with them. Of particular concern was the risk of ADM projects being perceived, developed and run primarily through the lens of being an IT project, without proper consideration of the fact that they are first and foremost about the lawful performance of an administrative and usually statutory function.

Our work over the last few years has really been about addressing those two concerns. Our November 2021 special report to Parliament titled *The new machinery of government: using machine technology in administrative decision-making* addressed that second point. We took a close look at how existing principles of administrative law and good decision-making apply when agencies seek to automate, including part automate, their administrative functions. The aim of that report was to get on the front foot and provide proactive guidance to agencies, with a view to helping them avoid maladministration, particularly before it came to us or some other body in the form of a complaint, investigation, royal commission or court action. The follow-up to that work was the report on Friday with the map of automated decision-making in the New South Wales public sector.

I will speak very briefly about that report. We engaged researchers at the ARC Centre of Excellence for Automated Decision-Making and Society, two of whom you've met in this hearing. It represents the first attempt in New South Wales to comprehensively identify and publish the ways in which the public sector is using or planning to use automated decision-making in the performance of their functions. The report includes a compendium of 275 such systems reported to be in use or planned to be in use. There are a range of interesting

and important observations made by the researchers in their report and a few takeaways to note. Automation is widespread, varied and increasing across both State Government and local councils. Currently, most of the use cases do not make a final decision and do not involve fully automated decision-making processes. Many of the systems reported through the project are currently not subject to any ADM- or AI-specific regulatory framework.

The research team looked more closely at a smaller sample of cases, and of those they found that less than half had had any legal input at the design stage. As I said, our focus in this work has really been about the here and now: what laws and principles currently govern ADM use by government agencies; how is ADM currently being used; and what can agencies do to avoid their ADM being found to involve maladministration. However, we also hope that our work will assist the Government, Parliament and the public to consider whether improvements are required to be made to the reporting and regulatory frameworks for ADM use in the public sector. Thank you.

The CHAIR: Thank you, Mr Miller. Ms McCallum?

RACHEL McCALLUM: Chair and members, I welcome the opportunity to appear before the Committee today. I would also like to acknowledge that we are meeting today on Gadigal land. As you may be aware, my appointment as Information Commissioner and CEO of the Information and Privacy Commission only commenced last week. I would like to thank now, therefore, the small but dedicated team at the IPC, as well as the Acting Privacy Commissioner, for their support during my first week. It is a challenging but also opportune time to assume responsibilities for regulating access to government information. As the new Information Commissioner, I am responsible for overseeing the transparency regime established under the Government Information (Public Access) Act 2009, known affectionately as the GIPA Act.

GIPAA's transparency regime creates a legal presumption that government-held information will be made available unless there is an overriding public interest against disclosure. As the IPC's submission to this inquiry notes, it the object of the GIPA Act to maintain and advance a system of responsible and representative democratic government that is open, accountable, fair and effective. A workable system of enforceable rights to access government information is, therefore, at the heart of New South Wales democracy. That transparency should continue to be protected as AI is deployed across the New South Wales public sector. As the Ombudsman has noted in his report on the adoption of automated decision-making, tabled last Friday, it demonstrates how technology in the public sphere is already being used at scale.

Since the IPC made its submission to this inquiry, there have also been significant international and Australian announcements about the future of AI regulation. In January, the Australian Government released its interim response to its consultation on safe and responsible AI in Australia. AI regulation in the EU and in the United States has moved forward. In November 2023, nation participants at the AI Safety Summit at Bletchley signed an historic declaration. All recognise the importance of privacy, transparency and explainability for managing the risks of AI. These developments can assist New South Wales to take a consistent approach to investment in transparency around AI at the sub-national level.

In the integrity space, such investment might include updating existing New South Wales access legislation to ensure it remains fit for purpose; expanding the remit of existing independent integrity bodies and supporting them to work together on AI-based issues; and providing additional AI governance support for agencies as they implement new policy or legal frameworks. The IPC's submission to this inquiry made a number of specific and important suggestions for refinements to the GIPA Act. The Act has worked well in its application to digital records, but these changes will ensure it aligns with increased technology-based decision-making by agencies. Our collective understanding of how to regulate effectively for AI, however, will evolve. I look forward, therefore, to working with the Parliament and the Government to develop useful responses now, without constraining our ability to adapt further in the future. Thank you.

The CHAIR: Thank you, Ms McCallum. I begin with a question to Mr Miller. The Government has developed the AI assurance framework. In your submission and your evidence just then, you were saying that machine technology, AI, ADM, whatever you want to call it—it sounds like it is already pervasive throughout agencies. Is that assurance framework robust? Is it being applied as you've seen it in the various iterations of the technology in our public institutions?

PAUL MILLER: The AI assurance framework we consider to be a useful step forward in the internal—within government—regulation of AI. There are couple of points to note about it, however. One is the limitations on its applications. Firstly, it is only prospectively applied. Many of the systems that were identified through the research report on Friday are systems that have been in place for some period of time, well before that assurance framework came into effect, so it would have no application to them. The other point is that the main oversight element of that assurance framework is reviewed by an assurance review board. That review only occurs in respect of a relatively small number of AI projects that fall within the criteria of the assurance framework. There is a

financial threshold—I think it is $5 million—and a definition threshold that it has to be AI as defined in that framework. Many of the ADM systems that were identified through the research probably would not fall within that technical definition of AI. It also covers projects that were funded through the Government's Digital Restart Fund, so—

The CHAIR: Sorry to interrupt. Do you think it is too narrow, then? Do you think it should be retrospective? I haven't read the entirety of that report. Do you think it should be retrospective and do you think that the thresholds are too narrow or too high, as it were?

PAUL MILLER: My understanding was that there was in development already within government an assurance framework 2.0, if you like. I caught Professor Oppermann's evidence, which seemed to suggest that there are question marks about the status of that. Briefly, to answer your question, I think the current version of the assurance framework is of limited value in terms of being a comprehensive regulatory tool for controlling the use of AI in automated decision-making in government, yes.

The CHAIR: In your report you said that AI is not operating in a legal vacuum. We already have administrative law that applies to AI as it comes. What areas of law reform do you think we need in terms of administrative law to make our systems more robust?

PAUL MILLER: It is a good question. Our focus as ombudsmen has tended to be on, effectively, compliance by government agencies with what exists now, rather than advising Parliament about what it should do about reforms in the future. What I will say is that administrative law is technologically agnostic. It sets out principles. It should apply, and it does apply, irrespective of the technology that is being used for decisions. But there will be some, at least, uncertainty initially in terms of how exactly it applies in some areas. I will give a couple of examples. One is the right to reasons. Not every administrative decision comes with a right to reasons, but some do. There are questions about what does that mean in the context of a decision that was made with the support of AI or other technology. Similarly, the right to be heard, which is a central element to procedural fairness—does your right to be heard entail the right to be told about how the decision was made? Because a right to be heard is, in effect, meaningless if you don't know what you're being heard about.

The other significant, I think, uncertainty gap is around the issue of bias. In administrative law, there is a rule against bias, but that rule has been developed around the concept of individual human beings having or being perceived to have bias in their individual decision-making. It is less clear that that rule would apply to the kind of systemic algorithmic bias that we're talking about as potentially occurring in the context of an ADM. So I think there is scope—and I heard Professor Bennett Moses speak earlier today about looking at particular areas of law, whether it's admin law, liability law or copyright law, and assessing the current legal frameworks in all of those areas to see whether there are uncertainties that can be made certain and gaps that can be filled. I think those are—some of them—in the administrative law domain.

Ms ABIGAIL BOYD: Good afternoon to all of you. Unfortunately, I have not yet read all of the report that was released on Friday, but I skimmed through some of it. Thank you very much for that detailed piece of work, which I think is going to be very useful. I notice the STMP—the suspect target management plan—comments, and I think this is really interesting because, as you say, this is a system that has been criticised for enabling intrusive and discriminatory policing. I was not aware that it had, sitting behind it, a method for assessing an individual's risk based on a new sort of AVR that is generating a shortlist of people to target. Can you talk us through how that happens without anybody knowing about it? And what should we be trying to do in response to that?

PAUL MILLER: Yes. Just specifically on the police example, one thing I should make clear is that our report is in four volumes. Two of those volumes are authored by our external researchers, and two of the volumes are authored by us. One thing we make clear is that the opinions expressed in the researchers' reports are their opinions and not necessarily the Ombudsman's opinions. I particularly need to make that clear in respect of police because, as you'll be aware, general oversight responsibility for police has moved away from the Ombudsman to another body—the LECC. Generally speaking, I don't consider it my place to comment or critique on police conduct. But to the broader point about authority, it's a really good one, because there's a point we made in our 2021 report which is that, under the rule of law, as an individual citizen you can do whatever you like unless it's prohibited. But if you are a government agency—an administrator—you can only do what you are authorised to do. So it's almost the reverse. Rather than agencies being free to do whatever they like unless it's prohibited, agencies like police can only do what they have been authorised to do and, in most cases, that means authorised by Parliament.

In the context of the adoption of technology, whether it's AI or what have you, there are very few pieces of legislation in New South Wales that expressly authorise agencies to use it. There are some examples. I think the Jury Act is an example where there's an express power for the Sheriff to adopt an automated decision-making

tool to select jurors—a good piece of legislation in my respectful view. But most agencies that are using technology are relying on what they believe is an implied authority in the legislation. So if the legislation confers a function on an agency—I won't use a police example—to make a decision, and that's all it does, then they would say that implicit within that is authority to use whatever technology assists in the making of that decision, as long as it's consistent with the legislation. I think there's a question for Parliament there, given that the authority ultimately comes from Parliament—whether it's express or implied, it's the authority given to agencies by Parliament—whether Parliament wants to, in some cases, make clear that that authority does not exist, so, whether there are use cases where Parliament says, "This is a use case we prohibit."

Even where that's not the case, if an agency is coming forward, particularly with new legislation, with a new power or a new function, I think it is open and appropriate for Parliament to be asking the question, "How do you intend to exercise this power? Do you intend to use technology?" and, if agencies are intending to use technology, to ask, "Why is that not made express in the legislation?" because, if it were made express in the legislation, Parliament would then have the opportunity to consider what safeguards need to be put around that technology—not just authorising it. But does that mean that we will authorise it but we will also, in the legislation, confer, for example, a clear right that, if the technology makes a decision with which someone disagrees, they have a right to an appeal mechanism through a human being? Should the legislation specify the degree of transparency around the technology? Should the legislation specify requirements around audits and testing that should happen with the technology—that sort of thing? Ultimately, to come back to your question, police are able to do this—assuming they are able to do it—because there is an implied authority under the legislation that confers their functions for them to use technology in the exercise of those functions.

Ms ABIGAIL BOYD: In your review, did you find any instances where the use of technology then needed to be approved by somebody at a higher level than the people using it?

PAUL MILLER: No, and this is interesting. I won't waste time by trying to find it in the research report, but there is a section of the detailed research report that talks about authority and where it came from, and perhaps one of the things that surprised me a little bit in the research was that point about at what level were decisions being made. There were very few decisions to develop and implement ADM systems that went to Cabinet, for example—that were identified as going to Cabinet. There were few that even went to the Minister, and it was apparent through the process of surveying agencies that it's not the case, for example, that the secretary of the department has visibility, let alone decision-making authority over all of the ADM systems within their own department. So it is certainly the case—and I think the research bears this out—that decisions are being made at a relatively disaggregated and potentially quite low level within an organisation.

Ms ABIGAIL BOYD: Another example, which I was referring to earlier today, was something I found out in police estimates a little while ago in relation to the use of facial recognition technology. The exchange I had there with the police officials was around "Have you tested this for bias?" The answer was no, they hadn't, but they didn't think that was a problem. One of the questions I was trying to tease out in one of the earlier sessions today was at what point is it okay that you have a biased system or a system that's potentially biased, where you then have a human in there as a cross-check before any action is taken in relation to an individual? I think we can see from the STMP that bias is not easily cured by having a human looking at it as well. Do you have any reflections on that? From your review, did you see that?

PAUL MILLER: I don't think there's anything in the review because the research was primarily about increasing transparency about the existence of these systems rather than seeking to audit, if you like, the systems against any of those matters. I would agree with your proposition that, if an ADM system is infected with algorithmic bias, having a human in the loop or human on top is not a solution to that problem. Part of the reason is the one you have identified, which is that the systemic bias may not be apparent and certainly won't be apparent in individual decisions that are going through that individual, which points to the other problem of algorithmic bias, which is that it also won't be apparent to the member of the public who is affected by decisions using the system.

That issue of, "Well, no-one's complaining about it," is also not a particularly rigorous response because of course no-one is complaining about it. The point about systemic problems with these systems is that they're not apparent at the individual level, which points to broader issues about administrative justice, because our administrative justice system by and large is built upon a foundation of if a decision is made that is wrong or unfair to you, then you have a mechanism, whether it's a judicial review, NCAT administrative decisions, a complaint to the Ombudsman. It's an inherently reactive system that relies on the individual to know that they have been wronged.

The CHAIR: Mr Miller, can I just interrupt there. It's a really important point, because you don't know what you don't know. You say in your report that we need, in effect, a registry for all of these so people can find

out where this technology is being applied and if it is being applied to them. How do you see that being enacted, if at all?

PAUL MILLER: I don't have a strong view about the calls that have been made from time to time about a centralised registry. The research that we did was not intended to be the starting point to create a registry and we certainly have no intention of maintaining it going forward. What I think, though, is important is that, for decisions that have a material impact on an individual, that individual, I believe, has the right to be informed if ADM was used in the making of that decision, an explanation that they can understand in general terms about how it was used, and the third element: the right to challenge that. The right to challenge that may not necessarily be a right to prove that the system was flawed or biased or what have you, because I don't think you can legitimately expect members of the public to do that. But it might be an avenue of, "I want my decision remade by a human being."

Ms ABIGAIL BOYD: Ms McCallum and Ms Minutillo, I might bring you in on this. Is there also an element of people assuming the computer is right? People assume that if a system has spat out that a person is of a particular risk of committing a crime or a person is not fit for a job, or whatever the system has told you, is it harder to challenge that because there is this kind of assumption that the computer must know better? Is that an element?

The CHAIR: A deference.

Ms ABIGAIL BOYD: A deference, yes.

PAUL MILLER: Yes, that is an issue. I think a related issue is the one of creeping complacency. AI, when used by government agencies, is part of a system and at the moment at least there are usually humans in that system as well and they're interacting. The issue with technological complacency is that, even if the system is designed at the very beginning to be lawfully compliant, reasonable, fair, with humans who are trained to understand the system, who are senior enough, competent enough, confident enough to form their own view and challenge the output of the machine—even if you get that right when the system is first launched—there is an inherent tendency towards complacency over time, whether it's the same people using the machine or in years new people are coming to use it and it's seen to be right most of the time. One of the challenges in this area—and I've talked elsewhere about the problem of talking about responsible AI as if it's a product that you just put out there, "Here's some responsible AI," and set and forget, and it's done—the problem can be that you've got to constantly be monitoring these sorts of things to avoid issues like the one you've raised.

RACHEL McCALLUM: If I can add to that, one of the ways in which it might be of assistance to ensure that there was that ongoing review by agencies or by the public institutions that are adopting AI is, as pointed out in the IPC submission, perhaps using the already well-understood GIPAA legislation to require additional transparency in relation to open access information of agencies and annual auditing by the IPC, for example—not of the technical capabilities of the technology, which may be an appropriate thing to have in another assurance framework, but in relation to enabling people and other institutions to understand what government is doing and to address any of the negative impacts that they may think are arising due to the deployment of AI. That transparency piece becomes a critical part of any future legislative framework. We have made some suggestions of not really significant rewrites of the GIPA Act but some additions that will perhaps support not only individuals who may wish to complain about a particular decision that was made in relation to them but also other civil society actors who may have an interest in questioning the use of technology by agencies.

The Hon. STEPHEN LAWRENCE: In terms of the policy frameworks around ADM applied in State Government, would you say that they're working in an adequate way or would you say there's a pressing need for legislative reform to set some principles around that?

RACHEL McCALLUM: Just in relation to access and transparency, the IPC's position is there is a need to start to address the aspects of access in relation to ADM or other forms of AI. So, yes, we haven't advocated for specific AI legislation. That obviously would be an option that this Committee and the Parliament and the Government will consider. But there is a need, as we point out in our submission, to address some of the anomalies that arise because of the way in which agencies are now, as the Ombudsman's report has shown, routinely adopting ADM to assist them.

In order to not undermine the access regime, which is so important in New South Wales, yes, we say that amendments may be necessary in a timely way and that that shouldn't prevent further amendments down the track. It shouldn't be seen as, "We must get this entirely complete at the first set of amendments." There may be others that would need to come and this is obviously an evolving space. We may find that there are other use cases or situations that require bespoke legislation, as I think the Ombudsman mentioned in his response earlier to Ms Boyd about particular enabling legislation might have its own special characteristics as it arises. But in relation to the

privacy legislation and the access legislation in New South Wales, yes, some amendments sooner seem appropriate.

SONIA MINUTILLO: Just to supplement that, we did in our submission talk to this concept of privacy impact assessments as being not currently mandatory under New South Wales privacy law but very relevant when personal information is being used as the dataset that is driving the use of these technologies to look at the risks, the harms and the mitigations that can be applied. In our submission, we talk to this idea that perhaps PIAs becoming a more formal part of the process, coupled with engagement with the Privacy Commissioner around those assessments, may also work to supplement existing frameworks.

The Hon. STEPHEN LAWRENCE: Have you got any thoughts, Ombudsman, on that question?

PAUL MILLER: Yes. I think there's a debate that, crudely put, is do we need legislation reform to deal with AI legislation? Do we need an AI Act like they have in the EU, or do we do a more bespoke kind of gap analysis in existing privacy law, information access law, administrative law or decision-making law? That's a big policy question. It's not really the job of an Ombudsman to be a policymaker, but I must say I'm a sceptic, for some of the reasons that you've heard from others before this Committee, about how we need an AI Act or what have you. But I do think that there are clearly gaps and anomalies in the existing legal frameworks—whether it's in the privacy legislation or under administrative law, whether at common law or legislative—where those gaps need to be addressed.

What I would be looking to do is kind of a gap analysis in all of those areas and saying that these laws were written at a time when this technology wasn't available. Yes, they are, generally speaking, all of them principles-based legislation. But does that mean that they're going to work exactly how we want them to in respect of AI/ADM? Algorithmic bias is a really good example of that, where I think the answer is that maybe it will. I don't know how a court will look at the issue of algorithmic bias, but are we going to wait until there's some terrible example to make its way through the court to find an answer to that? I think that there's definitely a call for reform in that area.

The Hon. STEPHEN LAWRENCE: Just out of interest, are you aware of any court cases in superior courts where procedural fairness and AI have been the subject of a ground of review, or something like that?

PAUL MILLER: I'm not, but can I take it on notice?

The Hon. STEPHEN LAWRENCE: Sure. Thank you.

The Hon. Dr SARAH KAINE: I have a general question. Thank you very much for your submissions and your evidence, and the most recent report that came out on Friday. Looking at all of that and looking at what your current roles are—and you've outlined that—how are your organisations respectively equipped for the changes that will be wrought on you by AI, both for the organisations that you regulate to some degree and for your own organisation? How are you dealing with what we're all facing, which is that ever-changing and evolving AI environment?

PAUL MILLER: Do you want me to start?

RACHEL McCALLUM: Sure.

PAUL MILLER: There are a few things there. The first is to recognise that we are an agency as well and that there are obvious benefits that are possible with this new technology, and that—as I heard the ICAC speaking about earlier—we should be considering the potential of the technologies in our own organisations. We've adopted a policy on generative AI use in our office which is quite conservative, I would say.

The Hon. Dr SARAH KAINE: Yes, I think I've got it here.

PAUL MILLER: The other thing is that we at the Ombudsman's Office have a very voluminous frontline service offering, so there's a lot of potential for AI in terms of better customer service and better customer experience in terms of tracking complaints, redirecting complaints et cetera. So there's that to start with. But in terms of our oversight function, two comments on that: The first one is about internal capability. If maladministration has occurred, whether it's occurred because a human has done the wrong thing or because—it's still a human doing the wrong thing, but a human has done the wrong thing by implementing this machine, the challenges of getting a basic understanding of the machine and how it works in the system is something that we are increasingly going to have to grapple with.

In the one investigation that we have undertaken in relation to ADM, which is the Revenue NSW one, the way we dealt with that was that we essentially worked very collaboratively with Revenue NSW to get them to describe—in language that we could understand—how the system worked, and that is not a complicated system at all. There will, I think, be a significant capability challenge for us. Related to that is the recognition that, where

we do investigate maladministration in respect of ADMs, it's quite likely, at least initially, that we're going to be coming to the party a bit later than we otherwise would, for all of the reasons that I discussed with Ms Boyd. So the need to be able to investigate quickly, because the systems are already operating—they're operating at speed; they're operating at scale. If they're doing harm, they're doing significant harm right now—so the need to be able to investigate perhaps more quickly.

The last point I will make in terms of investigation is that it would be—how can I put it?—easier. It would be easier, as an oversight body, if there were clear standards—and the AI Assurance Framework is a step in that direction—that agencies were required to comply with. For example, if there was a clear requirement that before any system was rolled out on the New South Wales public it had to be independently subject to an algorithmic bias test, my job would be easier because, if I get complaints, the first thing I can do is say, "Was it subject to that test?" If it wasn't subject to that test, that's maladministration, which is a much easier investigatory job than having to go, "Well, is it actually infected by algorithmic bias or not?" I look forward to the day when there are much more comprehensive and rigorous standards applied. Because that, essentially, is what I'll be oversighting, rather than the substance of the system.

SONIA MINUTILLO: From a privacy perspective, it is, as the Ombudsman has just indicated, an ongoing learning journey. We've done a lot of work in providing advice on a range of technologies as part of the Digital Restart Fund. That has enabled us to understand the types of technology by virtue of the types of materials that are presented as part of those projects and the business cases. But, alongside of that, the undertaking of those privacy impact assessments—which really map out what the information holdings are, what personal information is in question, where it's flowing, who has got access—allow us to actually apply that lens, from a privacy perspective, to what has been occurring and intersecting with the technology. But, equally, as the Ombudsman said, it sometimes takes engagement with agencies who are the technical experts to really communicate how the technology is in plain English, if I might put it that way.

The CHAIR: It looks like we don't have any more questions. Thank you very much for your answers today, which were very useful, and for the submissions you've made. There may be some more questions, especially once we've read that report more comprehensively, which we will get to you via the secretariat, on notice.

(The witnesses withdrew.)

(Short adjournment)

Ms LAURA CHRISTIE, Deputy Secretary, Digital.NSW, and Government Chief Information and Digital Officer, NSW Department of Customer Service, affirmed and examined

Mr DANIEL ROELINK, Director, Enterprise Architecture, NSW Department of Customer Service, affirmed and examined

Mr MARTIN GRAHAM, Deputy Secretary, Teaching, Learning and Student Wellbeing, NSW Department of Education, affirmed and examined

Ms JESSICA HO, Director, Digital Investment and Assurance, NSW Department of Customer Service, affirmed and examined

Dr ZORAN BOLEVICH, Chief Executive, eHealth NSW, and Chief Information Officer, NSW Health, affirmed and examined

Adjunct Professor JEAN-FREDERIC LEVESQUE, Deputy Secretary, Clinical Innovation and Research, NSW Health, affirmed and examined

The CHAIR: Good afternoon to our new witnesses. Welcome to the Portfolio Committee No. 1 inquiry and hearings into artificial intelligence. Do any of you have any introductory remarks or presentations you want to make before we start questions?

LAURA CHRISTIE: I have short statement if you're happy for me to read it.

The CHAIR: That would be fantastic.

LAURA CHRISTIE: First of all, thank you very much for having us here today. My name is Laura Christie. As I said, I'm Deputy Secretary of Digital.NSW. AI is a transformative technology that uses and learns from data to make predictions that can solve complex problems and inform decision-making. The New South Wales Government is taking a risk-based approach to support our teams and businesses in using data and technology ethically and responsibly to inform this decision-making. The key component of the New South Wales Government approach to AI is to build public trust that AI technologies are being used and developed ethically and responsibly and with a clear focus on community outcomes.

My role in the Department of Customer Service involves all-of-government digital strategy, investment and assurance as well as ICT and digital sourcing, amongst other responsibilities. In late 2023 the responsibility for the artificial intelligence program, policy and governance work was transferred to Digital.NSW from another division within the Department of Customer Service. We are leading a program of work to respond to the increased demand for AI and other emerging technology, including the need for all-of-government oversight of the increased risks associated with this technology. This program of work was considered by the secretary's board in June 2023.

DCS has observed significant growth of investment in and organic adoption of AI solutions across New South Wales Government. Leveraging these technologies presents opportunities to drive innovation in service delivery, productivity, streamlining processes and delivering more personalised government services. While this new wave of technology has many benefits, AI systems have risks and can have unintentional negative impacts if left without safeguards. Risks include the potential for AI systems to perpetuate biases, spread misinformation and cause privacy breaches, as well as legal and ethical risks when used in significant decision-making processes that lack human oversight. Advances in generative AI will provide further opportunities and risk.

To maintain leadership in this space, the New South Wales Government is committed to the values of transparency, community benefit, fairness, privacy, security and accountability and have enshrined these in an AI Ethics Policy. We are the first Australian jurisdiction to develop and implement a holistic and ethical approach to use of AI in the public sector. The New South Wales AI Assurance Framework came into effect in March 2022 and exists alongside the New South Wales AI Strategy and AI Ethics Policy as a key component of the New South Wales Government approach to AI. This integrated AI policy ecosystem supports public accountability over the use of AI and, we hope, helps to reduce public concerns over unintended misuse.

The AI Assurance Framework provides an umbrella framework promoting consistent AI risk management by all agencies. The AI Assurance Framework is in the process of being updated to integrate the elevated risks associated with generative AI and to integrate into the overall ICT Assurance Framework that is administered in my team and applies to projects over $5 million and all projects funded via the Digital Restart Fund. This means we can factor AI risks into the decision-making matrices of project assessments and investment decisions made by government and ensure that these risks are being treated in a cohesive way. It is anticipated

that the updated AI Assurance Framework released in late 2024.[1] We are also working on a guideline to support that AI Assurance Framework to make it easier for agencies to adopt the self-assessment framework into their governance, risk and compliance frameworks. Both will be available, as I said, in coming months.

In addition to this framework, Digital.NSW has released several public AI guidance notes on topics including end user guidance, simple terms and definitions of AI, cybersecurity and basic prompt guidance. We are also developing procurement guidance for New South Wales agencies considering the procurement of generative AI in one or more of its various forms, which will be available by mid-2024. New South Wales has also been co-leading the development of the national framework for AI assurance with the Commonwealth and will continue to contribute to its development by exploring how elements of the New South Wales AI Assurance Framework can be utilised, while ensuring alignment with the diverse needs of all jurisdictions. New South Wales collaborates significantly with its interjurisdictional State and Territory colleagues on progressing this vital body of work.

The CHAIR: Thank you, Ms Christie. Are there introductory remarks from the Department of Education or Health?

MARTIN GRAHAM: No.

JEAN-FREDERIC LEVESQUE: No.

The CHAIR: We will now turn to questions.

The Hon. JACQUI MUNRO: Thanks for coming, for the submission and for all the work that you do in this space. It's really important. I'm wondering about the leadership of this kind of policy development now that we don't have, for example, a Chief Data Scientist and the committee that was built around a range of experts from industry. Who is taking on that role now? How is guidance provided?

LAURA CHRISTIE: That work is being absorbed within Digital.NSW. We're in the process of reappointing the committee. We've reviewed its membership, as is appropriate, every two years or so. We're in the process of reviewing that committee. The committee will remain. As I introduced in my introductory statement, we're in the process of embedding the AI Assurance Framework into our overall ICT assurance. It's not treated as a separate issue to be considered; it's part of the overall assessment of ICT and digital investment, which we think will really make sure that it's embedded in the ecosystem of government. That's how that work is being taken forward.

The Hon. JACQUI MUNRO: The Ombudsman raised concerns that the framework wasn't necessarily embedded very well at this point. I'm wondering if the work that you're doing is in direct response to understanding that that has been the case and how you found that out.

LAURA CHRISTIE: Yes, we are working to embed the assurance framework because we wanted to make sure and be confident that it was being considered and used well by agencies. We think of the assurance team that runs this overall assurance of all programs over $5 million and all projects funded by the DRF as the front door to assurance, so we thought it made sense to combine those two doors—come in, register your project under ICT assurance and consider those AI risks as well. So, yes, absolutely that has gone into our thinking.

The Hon. JACQUI MUNRO: Does that mean that you need a bigger team to conduct that kind of work? It sounds like a pretty huge job.

LAURA CHRISTIE: We are resourced pretty well for assurance at the moment. But, absolutely, we are considering our resourcing arrangements within the Department of Customer Service and we'll be making an assessment as we understand the volume of work that comes in once we have gotten that approval to embed the AI Assurance Framework into the overall ICT Assurance Framework.

The Hon. JACQUI MUNRO: I'm sure you're aware of the Ombudsman report. They have this compendium of all the different types of ADM technology that is being used. Are these technology tools being developed within departments? Or are they being sourced from external providers at this point?

LAURA CHRISTIE: I can't talk about the specifics of every department. The role of Digital.NSW is to support agencies in their use of technology. We provide the assurance service. We drive digital investment and strategy. We administer the Digital Restart Fund. But, absolutely, I'm sure there would be a mix of both.

---

[1] In correspondence to the committee received on 10 April 2024, NSW Department of Customer Service provided a clarification to their evidence.

The Hon. JACQUI MUNRO: Perhaps that's a question for other members of the panel?

MARTIN GRAHAM: Everything we do in AI is under the New South Wales Government assurance framework. We work very closely to make sure that we are lined up because it's a rapidly emerging area, and we want to make sure that we're all under the same umbrella.

The Hon. JACQUI MUNRO: But they're not developed in-house? They're sourced externally and then made to work appropriately? I'm not saying that's a bad thing necessarily. I'm just curious to understand how those technologies are being adopted.

MARTIN GRAHAM: Maybe when we get into Education I can explain how that works.

The Hon. JACQUI MUNRO: Sure.

ZORAN BOLEVICH: In Health's case, we have examples of both. We have examples of commercially based products that have been acquired and are going through a process of systematic adoption—again, through the use of the framework as well as other additional considerations that we have in health care, which we might touch on later. There is a structured process of adopting those third-party technologies into our environment. There are also examples of original work, either through research projects—arising within our health system, we have a significant research output that Professor Levesque might want to talk about later—but also digital innovation that's occurring in NSW Health. So we have examples of some home-grown, early AI initiative, let's call it that.

The Hon. JACQUI MUNRO: Are you sharing those between departments at this time? Or do you think that the use cases are generally so specific that there's not much value in other departments having an understanding of where other departments are?

MARTIN GRAHAM: I think there's a really great value. One of the valuable parts of this process that has been brought together was looking at how NSW Health are planning to even just log the use cases, so certainly that's something Education will be looking at. At the moment, there's not a lot of overlap, because we're fairly specific in what we do. But as this grows, there'll definitely be areas which are very similar. It might be in admin workload reduction, which I know we're all passionate about trying to do, and having a central point across government that helps draw us together is really helpful for that.

JEAN-FREDERIC LEVESQUE: In the space of research and innovation, there is obviously a lot of collaboration already with industry and trade. That includes a lot of advances using artificial intelligence for clinical trials or in biotech manufacturing. We're sharing not just the actual outcomes but we're collaborating in developing a lot of those solutions.

The Hon. JACQUI MUNRO: Finally, we've heard a little about regulatory sandboxing or at least the opportunity to essentially run experiments on different technologies so that they can be proven before released on a mass scale. Do any of you have examples of having done that and could you please elaborate?

MARTIN GRAHAM: Certainly. I think NSW EduChat would fit that criteria. For Education, we fit under the Government's total AI Assurance Framework, but then we also have—I think you heard about it this morning—the National Australian Framework for Generative AI in Schools, which is something that, certainly, the New South Wales Minister took a lead on. What that does is it makes it a bit more specific around education—those privacy issues, ethical issues, and, really, the use of AI as a tool for teachers. Our own product is NSW EduChat. We're currently trialling that in 16 schools. It's a solution that is really designed to address those kind of issues in the framework. It might be helpful if I went through the key features of it. It is one of the best way of understanding how it addresses all those risks.

The first thing is that the data is maintained in our systems. It provides us with—and you were talking about—that kind of sandbox or the walled garden environment so that the users don't have to all be constantly monitoring what they're doing because we have provided that safe environment for them. It also makes sure that it gives teachers answers that are suited to their role and to the fact they're in New South Wales. It's got what is known as a system prompt so that they don't have to go through each time and type in, "I'm a New South Wales public schoolteacher. This is my context." It looks at the New South Wales syllabus and so on first, so it's not drawing on international examples, which we probably all have experience of. It's got what's called an orchestrator, which is probably one of the most useful parts of it, which means that when someone types in a question, something they want to work on, it will decide which of the AI models will be the most effective for it. You will be aware that some do maths better than others, some do different tasks. It will actually direct, whether it's a teacher or a student, to that particular model so they will be able to get a more accurate answer.

It's got what's delightfully described as constitutional alignment. It goes and checks answers against the department's values. It actually has this much deeper look. It also has, for students, jailbreak prevention so that they can't—you know, give a teenager something and they're going to try to break it or make it do something that

you don't want it to do. It has that built-in profanity filtering. You're not going to be able to swear, and it's not going to swear at you. Most importantly for us, really trialling—when you type in, "Write me an essay on Hamlet," it's not going to write you an essay on Hamlet. It's going to ask you, "What do you know about the book? What can you tell me about the characters?" It kind of takes you through that process as well. That's something that we have rolling out in 16 schools at the moment to really learn. It meets all of the New South Wales assurance framework. It meets all of the national requirements. It gives us a chance for students and teachers to go and play with that in that safe environment.

The Hon. Dr SARAH KAINE: Thank you all for appearing and for the submission. A lot of what we're talking about is about regulation. We heard a lot about different—or there seems to be a bit of a consensus around law reform. Ms Christie, and maybe Mr Roelink, I just wondered about your thoughts at the operational end. I don't know if you've had a chance to hear any of what's come before. I'm sure you're well aware of the approaches. I just want to get your thoughts on regulation, what you feel is appropriate or where the boundary should be, and the different approaches.

LAURA CHRISTIE: Thank you. Yes, absolutely I think there's an opportunity for New South Wales Government in this space. I'm very conscious that the Commonwealth is currently considering legislative change around AI, but that doesn't foreshadow New South Wales also considering changes in the New South Wales legislative and regulatory environment. My team is currently considering some work by the James Martin institute that they released last year that really talked to what a subnational regulatory framework would look like for AI, and that's under consideration. But we're also very conscious that we need to, if not be led by the Commonwealth, be consistent with the Commonwealth in the direction that they're heading and understand where they're heading in the first instance as well.

The Hon. Dr SARAH KAINE: I note, and you talk about it in your submission, the ethics principles underpinning the current policy. My concern is not that there are ethics principles, which is obviously a good thing, but without any enforcement or any body that is checking whether that's being applied, isn't there a danger that that's just a nice set of principles that doesn't actually impact on how this operates in practice?

LAURA CHRISTIE: Yes, sure. I think from the centre of government perspective—and we've sort of lived this experience in cybersecurity as well—we've certainly made a conscious decision that our role is to support agencies in their use of AI rather than kind of mark the homework. We want to help them do their homework well, is kind of an analogy.

The Hon. Dr SARAH KAINE: But is there anyone marking the homework?

LAURA CHRISTIE: The Auditor-General absolutely would be. The responsibility for the use of AI ultimately rests with the agencies as business owners, and the CIOs, the CISOs and ultimately the business owners who are using AI in their own business. Our role from the centre is to support them on the best way that they can do that. We've released a number of guidelines to support public servants in the use of generative AI and we're updating the AI assurance for the risks of generative AI, as two examples of that. We do also think the more prevalent use and understanding of that AI Assurance Framework will lift capabilities. But I think that's not to say that if we saw significant harms or examples where the assurance framework had not been used and we were getting problems as a result, that we wouldn't consider a more significant intervention in that space.

The CHAIR: The threshold and boundary for the assurance framework, the updated one, is $5 million of investment or any funding under Digital Restart. Why is that the parameter? Does that assurance capture local government and investments in local government? We've heard that it is being adopted by decision-makers at a local government level. Could you respond to that?

LAURA CHRISTIE: Yes. I probably should clarify, the AI Assurance Framework was mandated itself. It was mandated by a circular—I think it was in May 2022.

DANIEL ROELINK: March.

LAURA CHRISTIE: March 2022, thank you. That applies to all New South Wales Government agencies. I couldn't talk to whether councils are using the AI Assurance Framework. But it is publicly available, so they could choose to consider the issues that it has raised as part of that.[2] What I'm talking about is that our intention with the $5 million is to embed that AI Assurance Framework in our overall assurance of ICT projects, which my team administers on behalf of Treasury. It's the gateway process for ICT investment. That threshold is

[2] In correspondence to the committee received on 10 April 2024, NSW Department of Customer Service provided a clarification to their evidence.

set at $5 million because it's kind of the appropriate level of kicking off an assurance—having the scrutiny on a project at that $5 million level and anything funded by the DRF, because we think, as I understand it, the scrutiny is needed on those projects as well from an assurance perspective.[3] But we're embedding the AI Assurance Framework because we want to make sure that all ICT investment is captured as a result.

DANIEL ROELINK: I will just add to that. The assurance framework also is to be applied across anything that's below $5 million, or an operational system. If there's a residual risk that can't be mitigated that is above median with the current framework, then they're recommended to attend the AI Review Committee, which has the experts. The latest version of that has a higher threshold, but there are additional controls that are put in place in the new version to mitigate risk, particularly around generative AI.

The CHAIR: What are the new risks around generative AI? What are the top three or four risks as you see them?

DANIEL ROELINK: They're fundamentally the same risks that relate to the ethical principles within the ethics policy that has been mandated for reference and use by New South Wales agencies. The lower-level detailed risk is really around the ability to be able to explain the decisions or the output from using that generative AI model. It's really a focus around assessing the risk of using that technology when you're unable to explain the basis on which the output has been generated through the algorithms and models, noting that those models are trained on billions of parameters and it's impossible to be able to describe and explain why you might be getting a certain outcome and output. With the guidelines we've released around generative AI use, we also do advise that agencies really consider the implications of using generative AI platforms. They are some of the guidelines we referenced in the opening statement in which we're again advising from the centre of everyone being careful about how you might be using this type of technology when you're unable to explain or prove where the data has been trained on through those large language models.

Ms ABIGAIL BOYD: Have there been any examples of uses of AI or ADR have that been turned down by the Government for not meeting tests of being ethical enough?

LAURA CHRISTIE: I recently inherited the AI ethical assurance framework, so I can't talk to the specifics of what the AI committee considered. Jessica, I don't know if you have any answer to that? But I would also say that it's a self-assessment tool as well. Agencies can self-assess their projects according to the framework. So, absolutely, I'm sure there have been examples where people have felt that the risks might not have outweighed the benefits and have considered that it might not be the most appropriate use of the technology.

JESSICA HO: Maybe I'll just give a bit of a generic what type of risk we actually see the projects coming up with. The type of risks that we've identified with those projects coming up to the AI Review Committee are in the form of data quality, data security, legislation adherence, use of personal information in the solution, right of data, and alternative pathway if the AI solution did not work. Those are some of the types of risks we have identified with those projects being reviewed by the AI committee in general.

Ms ABIGAIL BOYD: The self-assessment is by the agency who is proposing to use it?

LAURA CHRISTIE: Yes.

Ms ABIGAIL BOYD: Is that something that is standard? I'm thinking about the Auditor-General recently looking at cybersecurity. That was all done on agency self-assessment, and it was found to be not particularly robust. Is there a reason why that's the case?

LAURA CHRISTIE: Ultimately the risk sits with the owners—the CIOs and the CSIOs and the business owners who are using the technology. The assurance framework has been set up to help them navigate those risks. If you were to, for example, centrally say, "You must come through this assurance framework that we will assess at the centre," with the prevalence of use cases of AI across the New South Wales public sector it would be a considerable bottleneck from the centre. It also would probably not support capability uplift about the use of these technologies in one's own business.

Ms ABIGAIL BOYD: The example I've been giving today, because it disturbed me when I asked the police officials about this during budget estimates—there was a particular facial recognition technology and an ADM that was using particular technology that was shown to have some racial bias in it in US examples. When I asked about it, they said that they hadn't done any bias testing. For the use of that, would that have just gone

[3] In correspondence to the committee received on 10 April 2024, NSW Department of Customer Service provided a clarification to their evidence.

through a self-assessment process where the police decided that they were going to use that, and they ticked it off themselves as being not risky?

LAURA CHRISTIE: As you work through the assurance framework—and we would be happy to table it for your consideration. It is mandated by a government circular, as I talked about. Agencies are required to assess their use of AI against the framework. And then there are triggers that push a project into medium or high risk that then goes to the AI Review Committee for consideration. I can't talk to the specifics of that example.

Ms ABIGAIL BOYD: Are these subjective factors? Are they things that the agency themselves have to look at and say, "Yeah, this probably pushes it up into the category where it needs to be looked at"?

LAURA CHRISTIE: I wouldn't say it is subjective; I would say it is against an objective framework. But, yes, agencies are making those assessments themselves.

Ms ABIGAIL BOYD: One of the things we found in the consultants inquiry which was causing a lot of issues was this reliance on agency self-assessment. It's all good and well to have rules and it's all good and well to have somebody saying that they're complying with the rules, but whether or not they're complying with them is a completely different story. Who is checking?

LAURA CHRISTIE: I would say to the earlier point about us integrating the AI Assurance Framework into our overall ICT Assurance Framework—that's going to give us a much better view into the use of AI in ICT investments and a much more transparent approach into seeing where agencies are using the AI and making sure that the framework is being utilised and applied.

Ms ABIGAIL BOYD: With something like that and with the STMP, where people have been raising concerns around privacy and over-policing and a bunch of other things—now that those things are in play, what happens if further concerns are raised? Where is the machinery of government? Who then can look and say, "Well, maybe that's a problem"? Or do we have to rely on the agency to take another look?

LAURA CHRISTIE: I'm not sure about the STMP that you've—

Ms ABIGAIL BOYD: Sorry, the Suspect Target Management Plan, which I've just seen today because it was in the Ombudsman's report that was tabled on Friday saying, "Actually, that's got this complicated AI sitting behind it that generates a list of anticipated targets." That is a program that has faced a lot of scrutiny for overly targeting vulnerable people, including First Nations people and people with disability. If we now are looking at that and going, "Well, hang on, we've got this better appreciation now of AI. We're going to have a look at that again", who is going to do that bit of work? Or are we relying on it to be uncovered and put in the media before we get someone to revisit that?

LAURA CHRISTIE: I would say that the role of Digital.NSW is not as a regulator. We support whole-of-government policy and capability uplift from the centre. There are a number of other ways that issues like that would be drawn out—before it hits the media, to your point. But that is not the role of Digital.NSW and the administration of the framework.

Ms ABIGAIL BOYD: To conclude on that then, so there is no-one within government that checks?

LAURA CHRISTIE: On the application of the AI ethical assurance framework?

Ms ABIGAIL BOYD: Yes, that checks that and also looks back and goes, "That probably shouldn't have been let through." These technologies can have enormous benefits, but they can also have enormous risks, particularly to the public. We may not have understood when we first put them in place, but now that we've got a better understanding, where in government sits the responsibility to then check and look and readdress those risks?

LAURA CHRISTIE: I would say that there's a role for the audit and risk committees of each of the individual agencies and a role for the Auditor-General in this space.

JEAN-FREDERIC LEVESQUE: Can I just say also that, in a sector like health, we do have a lot of clinical governance processes in place that constantly review medical devices that may have algorithms based on artificial intelligence. They're evaluated. We maintain a living evidence review because it's a field that's emerging very quickly. We've got a new technology assessment unit that also revises those things. It's embedded into the clinical governance, obviously, because any intervention that changes the way care is delivered is appropriately evaluated and has to be from a safety and quality perspective.

Ms ABIGAIL BOYD: It depends on the department.

JEAN-FREDERIC LEVESQUE: I just wanted to say that, from a Health perspective, we do have processes in place for that.

The Hon. CHRIS RATH: I wanted to move from the risk side of things more to the positive side of things. I was wondering if each of the departments could give some good examples of how AI technology is already being used for government service delivery or what you see are some potential great examples that you're working on for the future as well.

MARTIN GRAHAM: We're only a few weeks into the NSW EduChat trial, but I had the privilege of talking to Tim Lloyd, who is the principal of Plumpton High School, one of the schools on the trial—a very active school. He's got a very engaged senior executive. He's actually been trialling it across each of the key learning areas. Teachers have been using it. Students have been using it. One thing they are finding already is there are a number of ways in which it helps that are kind of quite context specific. One of the really low-value things teachers do is just every day giving kids instructions and things. One of the little examples that popped up was a languages teacher had to tell all the kids, with all their different types of computers, how to change their keyboard to simplified Chinese—one of those things that is just a real irritation for teachers. They thought, "I'll use this new tool I've been given." Typed it in—"Look, can you give kids the instructions for how to do this." It came up with a very simple set of instructions on the interactive whiteboard and the kids managed to do it in a very short period of time.

We're really interested in some of the low value stuff, so the equivalent of primary schoolteachers cutting things out and spending all weekend pasting it in. That is all now—the equivalent electronic being able to do that really frees up teachers. They've often got a bachelor's degree and a two-year master's degree and we've got them cutting things out. It means they can apply their minds to that high-level work. Because we want it to be a really teacher-directed tool. It really amplifies what they can do in the classroom rather than replacing them. So those kinds of things—different staffrooms have been doing different things with it. They're using it jointly to undertake helping them with preparing materials for classes. Tim was also talking about the amount of time it saved him. If he wants to teach a class on an engineering topic, it can actually help him get those materials together. But in a way that—one of the key assurance frameworks is that humans are always in charge. You're always responsible for the output of the model. It is enabling them to be able to exercise that. But it's early days. One of the key things we're doing is getting it out there amongst teachers and students to be able to see what they come up with in this safe environment.

The CHAIR: Further to that, Mr Graham, the Committee visited UNSW's facility and looked at their quantum computing and they have an amazing VR classroom facility there. Within that, students—and these are high-end engineering students—were able to access a virtual robotics machine, robotics equipment that, if they wanted to use in real life, would cost a million dollars or thereabouts and they would get access to it once a year. But because they had simulated it in virtual reality, the students could use it all the time. Does the department have any plans to do a similar thing—not virtual teachers, but access especially to virtual classrooms, virtual materials?

MARTIN GRAHAM: I think that's an example of how AI is just expanding so quickly across every field that we'll be tapping into—eventually it's something that will be ubiquitous and we'll be able to take advantage of it. We're not currently spending a large amount of money like UNSW might be on developing that kind of technology, but we're certainly interested in adapting and adopting it when it's available. We're a large education system but, compared to the rest of the world, we've probably got a relatively modest research budget in that regard. But particularly for rural and remote kids, already they are the number one beneficiary of the ability we have as a system to be able to provide classes across distance. I think since COVID it has got a bit more sophisticated and we've been able to—culturally kids understand it. We've got the tools and technology to do it. We certainly see AI is going to be part of this technology. It's going to be part of every bit of technology. If it can help us with those things, we'll certainly be doing it.

The Hon. JACQUI MUNRO: How are you finding the places that AI can help? What's the ground-level, frontline approach?

MARTIN GRAHAM: The trial is certainly part of that, because it's moving so quickly. I think very rightly as a system we're making sure that it's all safe and dealing with all those things, but we're not going to be able to find all the use cases. Teachers are out there. They're the professionals on the ground. They understand their content. They're excited by it. We've got this huge kind of groundswell of enthusiasm, so we're keeping it safe whilst we learn from them. That's what we're doing. We're not doing a three-year program. It is terms 1 and 2 and we're going to rapidly look at our experience from that, the data that comes in and the richness of their experience as well.

LAURA CHRISTIE: I would add to that. The Digital Restart Fund has recently released a series of priorities and called for projects to fund. One of those priority areas is AI to support frontline systems to reduce administration of frontline workers specifically, so that's currently being considered.[4]

The Hon. JACQUI MUNRO: So other departments are speaking to staff and asking for ideas about how we can help make your lives easier and more efficient?

ZORAN BOLEVICH: Yes. Speaking from Health's perspective, like many organisations we've started our journey in the back office of Health, looking at how we can improve productivity and cost efficiency of things like ordering, procurement, supply chain management—those kinds of things. Then we moved to areas where we can reduce administrative burden for our busy frontline clinicians. A very good example of that is dictation systems that use natural language processing. These days our radiologists right around the State use that kind of technology to dictate their reports. That saves a lot of time and cost. Then the natural progression from that is focused on safety-improving technologies. Health is a very safety-conscious industry. We are constantly striving to avoid and reduce avoidable patient harm. A good example of that is a trial we've got of the use of AI to support early detection of sepsis, which is a life-threatening condition in hospitals.

The Hon. JACQUI MUNRO: I did hear about this trial, actually.

ZORAN BOLEVICH: And then a bit of a combination of both productivity and quality with a couple of really good projects, one led by our colleagues in the Cancer Institute NSW. They're currently working through a process of trialling some deep learning technology that will help them with interpretation of mammograms for their breast screening program because the demand for that type of service, and then the radiologists who can support them, is ever increasing. They're looking at technologies that aid in managing that workload. Another example is one of our districts that is using image recognition technology to support the wound care management, including in the community. So we're starting with the productivity, moving into safety and quality and increasingly moving into those more clinically orientated applications. How we find them—I think Jean-Frederic can answer that question, as we've just formed an AI taskforce to help us do that.

JEAN-FREDERIC LEVESQUE: In addition to what Dr Bolevich has just said, we also are proactive in working with academic institutions to explore new AI applications that will help to accelerate research in drug discovery, for example. We've supported programs in finding new markers of diseases. So both in terms of new diagnostics as well as new ways to treat patients, we're proactively funding and partnering with academic institutes in integrating AI, and also in the data and analytics space, where we know we can free up analyst time by having machine learning applied to actual data assets. I will say that the last aspect is in the space of laboratory medicine where our organisation, NSW Health Pathology, for example, is actively seeking to use AI as a way to improve that diagnostic testing accuracy but also processes and management of samples and the way we analyse them. We've got a few different areas where we systematically review the literature so that we capture what's happening internationally and embed or explore those AI-based technologies that are demonstrating effectiveness in studies. We also work proactively with our own expertise within New South Wales because we do have quite strong teams that can embed AI in many aspects of research and innovation as well.

The CHAIR: The State Government and all its agencies is a massive repository of data. When you're working with a research institution on an area in health, with all the privacy concerns, how do you make sure that that data is handled mindful of those privacy concerns and kept as a discrete packet of information or data and not ingested into another model somewhere else? How does Health manage that? Do they have a framework for managing these types of risk?

JEAN-FREDERIC LEVESQUE: It's exactly like for any other conduct of research or any conduct of using of data for the purpose of managing the healthcare system, so the same rules apply. We don't allow data to be sent abroad as part of any packages. We increasingly use also what we call secure interfaces that enable people to analyse data without extracting the data. That prevents concentration of data in systems that we wouldn't be able to control afterwards. The same data governance processes apply regardless of if it's standard analytic methods or AI-based analytic methods that are used. We're looking into this through our AI taskforce that Dr Bolevich and myself are co-chairing. We've got a specific group that is looking at the data governance policies. We're going to make sure we update them as AI technologies emerge and enable more capabilities so our policy is always following those developments.

[4] In correspondence to the committee received on 10 April 2024, NSW Department of Customer Service provided a clarification to their evidence.

The Hon. Dr SARAH KAINE: Associate Professor Levesque, you mentioned this AI group but there was also a mention of you setting up a technology assessment unit. Are they two distinct things?

JEAN-FREDERIC LEVESQUE: Yes. That technology assessment unit is an ongoing unit that has been there for many years within New South Wales. They have a specific role with regard to this, especially when medical devices or other technologies that we want to introduce into Health have an AI component in them. From that perspective they would evaluate and assess that. They're not in charge of evaluating all of the AI digital platforms—Zoran has got a team and a process that does that—but when AI is integrated into another therapeutic technology, the new technology assessment unit has a role to play.

The Hon. Dr SARAH KAINE: So it is for therapeutic technologies rather than overall?

JEAN-FREDERIC LEVESQUE: Yes.

The CHAIR: Thank you very much for all your work and for making the time today to provide such useful evidence. The Committee and I really appreciate it. If there are any follow-up questions, the secretariat will be in contact in due course to provide those to you. All the best with all your endeavours.

(The witnesses withdrew.)

The Committee adjourned at 16:35.