Journal & Proceedings of the Royal Society of New South Wales, vol. 159, part 1, 2026, pp. 127–142. ISSN 0035-9173/26/010127-16

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

Panel Session 5: AI — Future Directions[1]

Hugh Durrant-Whyte, NSW Chief Scientist & Engineer
Jie Lu, University of Technology Sydney
Pascal Van Hentenryck, Georgia Institute of Technology
Toby Walsh, UNSW Sydney
Lina Yao, CSIRO Data61

Sally Cripps: Our final panel for today is AI and future research. It's chaired by the NSW Chief Scientist and Engineer. He's going to give some closing remarks at the end of the panel.

Hugh Durrant-Whyte

We have the world's best people for what's going to happen in the future, which is a great subject to end the day on. We have Jie Lu from the University of Sydney; Pascal van Hentenryck, who gave the Keynote, from Georgia Tech; Toby Walsh from UNSW; and Lina Yao, who's currently at CSIRO, but is about to move to UNSW as a future fellow. I'd like to talk about the future of AI. There are lots of diverse views, and I'd like to do it in two parts. First, I'd like to talk about what's going to happen in the next one or two years. What is the current thing going to look like? Should we all be out investing in AI shares or should we all be careful and buy a bunker and put it in our garden somewhere? Second, what should we be doing in more like 20 or 30 years? How will it change humanity? I think that's a different take on these things. Jie Lu, I'll start with you.

Jie Lu

I'm Jie Lu from University of Technology Sydney. Currently, I have two roles. First, I'm the director of the Australia AI Institute at UTS. We have 30 staff members covering almost all areas of AI: machine learning, computer vision, natural-language processing, and robots and so on. And 15 postdocs, and 220 PhD students. Quite a large AI institute. In the last four years, UTS has been ranked top in AI in Australia and one of top 10 globally. Second, currently I'm an ARC Laureate Fellow; my topic is autonomous machine learning to support decision making in complex situations. This will be finished early next year and immediately I will move to my industry Laureate Fellow to work on AI for women's health.

The first question: one or two years. That's relatively easier than 20 years. In one or two years we will have some directions, particularly from a research point of view. The first direction is autonomous AI systems. You have lots of apps in your computer, on your mobile phone. What we are doing now is AI agents. An AI agent system will deal with the tasks, and close all the apps to minimise human supervision. At the same time, AI with autonomous machine learning will

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

autonomously pick up relevant data to support human decision-making. The second direction is AI for science. Science has lots of unsolved problems: in mathematics, physics, chemistry, and health and so on. AI will drive new science discoveries. Two very important directions, I think.

Hugh Durrant-Whyte: I certainly agree about the science one. Last year, two machine-learning researchers won the Nobel Prize in chemistry. That shows you what's changing in scientific discovery. Lina, give me a brief introduction and what do you think about the next two years? What are the big problems?

Lina Yao

I'm Lina Yao. I was with CSIRO Data61. I was a senior principal research scientist and science leader for translational machine learning. Very soon I will move to UNSW as a professor and carry out my ARC Future Fellow project. My background is machine learning and data mining.[2] We wrap up everything as AI. My research has been gradually evolving from future learning, generative modelling, and reinforcement learning, to deep learning. Currently, my focus is lifelong agentic AI for multimodal decision-making. I will focus on addressing three critical research challenges. The first is alignment: it's a multifaceted concept to align the LLM agent with human values, as well as physical laws, for example, to handle some knowledge-intensive tasks. The second focus is how to equip the agent with the capability to self-involve by learning from the interactions and the dynamic and contextual environment. The third focus is how to further improve reasoning ability.

In the next one or two years, I believe we will witness the thriving of AI agents, especially LLM-powered agents. I can summarise it in two words: first, *verticalisation*, because everyone has been making great efforts to transfer the general ability of LLM agents to adapt some domain-specific tasks. That means how to transfer the ability from a general purpose to vertical domains like eBusiness or other things. The second word is *personalisation*. The agent will involve and learn to customise itself to fit each individual. That will be of great benefit, especially for education and e-commerce.

Hugh Durrant-Whyte: Toby, you've been in the area quite a while now. I guess you've seen ups and downs in the AI area. I'm really keen to get your view as to where you think things will go in the next couple of years. Are we going to see huge growth? Are we going to end up with a fall? And a bit of background, too.

Toby Walsh

A little bit of background is that Hugh used to be my boss when he was CEO of NICTA[3] and Pascal used to be my boss when he was a group leader at Data61. I'm pleased to be on a panel with my ex-bosses who still want to talk to me. I think it's easier to predict 20 years in the future than a few years in the future. I think it's easy to say we'll be using AI to accelerate the way we do science, transforming the way we do science, to do things that look at data sets that humans just don't have the capacity to look at, in order

2 Hawkes H (2021) This award-winning engineer turns big data into useful AI applications. *Create Digital*. https://createdigital.org.au/engineer-wants-to-help-ai-become-more-human/ [Ed.]

3 See https://en.wikipedia.org/wiki/NICTA [Ed.]

to transform genetics and medicine. In 20 years' time, you can be confident that we'll have completely transformed the nature of those pursuits.

It's much more difficult to predict the next couple of years. Until recently I thought that the AI bubble was yet to burst because we are only now talking about it being the "AI bubble." I remember we were talking about the dotcom bubble for several years, in the early 2000s, and it was only after we stopped saying it was a bubble — but maybe this is a new type of economics — that the bubble really burst. I was thinking maybe since we've only just started talking about the AI bubble it might last. But then yesterday I read that Google is proposing putting AI data centres in space. It's madness. Why would you spend money trying to get GPUs up into space? It really cannot make any sense other than the fact that you can.

Hugh Durrant-Whyte: 24-hour-a-day solar panels?

Toby Walsh: Well, yes, but we're a vast continent. The sun shines most of the time. We have batteries as well if we want solar in the middle of the night. It sounds to me like the worst excess is to think that we should waste the planet's resources by sending rockets up into space to have data centres up there. But let's move on.

It's unprecedented that we're spending billions of dollars a day, the scale of investment happening. Clearly some of this is very speculative. A number that caught my attention the other day: 2.5% of the USA's GDP is being spent on AI data centres. That is almost again without precedent; there's only one precedent I could find, which was at the height of the nineteenth-century railway boom: when they were building the railways and the industrial revolution in the US, they were spending two and a half percent of GDP.

But there's a useful historical lesson to be learned there: lots of people lost money. There were lots of public bonds being invested in the railways. A few people — the Vanderbilts and the Rockefellers — made a lot of money in building the railways, but most people lost everything. But building the railroads laid the infrastructure for the economy that became the United States. The same I think is true here: the bubble will continue to expand for a couple of years; maybe in two or three years' time, it will burst. A lot of people will lose a lot of money. A few people will make a huge amount of money, but that's irrelevant in the long-term scheme of things because that's laying the infrastructure for the new type of economy — the AI-powered economy — that we will have in 10 or 20 years' time.

Pascal Van Hentenryck

I'm going to disagree with you guys. I think it's much easier to predict the next two years than the next 20 years. I'm going to tell you why. Toby started stealing what I wanted to talk about: the dotcom bubble. What happened in the dotcom bubble was very interesting. It came to the public consciousness that there was this thing called the Internet, and everybody started building websites and e-commerce and so on. 99% of them failed. One of them wildly succeeded and is now one of the largest corporations in the world. So this is what's going to happen in the next two years: everybody's going to build an AI agent for X. Everybody's going to build a chatbot for X, and 99% are going to fail, but some of them are going to succeed and do something very interesting.

We'll talk again in two years, but I think that's going to happen.

Toby Walsh: Yes. You're right. Many of those businesses failed in the dotcom bubble, but two — Webvan online groceries and pets.com — are now proper businesses. Those things became real businesses but it took a while.

Pascal Van Hentenryck: That's the first thing. And there is another thing, a big one. If you've seen eight-year-old girls playing soccer, they bunch together all the time. They just don't know how to play. This is going to happen in science. Everybody is saying, oh, LLM is the new tool. I'm going to do LLM for X, for Y, for Z, for everything. All scientists are going to do that. It's obvious: 99% is going to be irrelevant. But some are going to do some interesting things. They're going to try to make them more trustworthy. That's what's going to happen.

And the third thing that's going to happen is that there will be a lot of good work in AI for engineering that is already getting close to being mature. I completely agree with Ros: academics are good for doing proof of concept, but you have to bring things to market. I think that's what's going to happen in the next two years. Finally, I'm going to agree with some of the things that you have said. What's going to happen in the next two years is AI for scientific discovery. The Nobel Prize in chemistry was an exception because they had massive amounts of data. What's going to happen in the next two years is accumulating a lot of data on many different scientific fields such that you can do the same thing in those field. I'll stop there.

Hugh Durrant-Whyte: Let me start by disagreeing with the last thing you said. One of the things that's interesting about that Nobel Prize is they had almost no data. They only had 200,000 protein sequences and they used it to predict 20 million of them.

Pascal Van Hentenryck: Yes, but they had secrecy. They used —

Hugh Durrant-Whyte: A good model.

Pascal Van Hentenryck: And I think that's important. And it was not an LLM, so this was very special.

Hugh Durrant-Whyte: I want to unpack that slightly before we move on to the long term. There is a big focus on LLMs. Like the engineer who said something a little earlier, I do have a concern that you cannot say or provide any guarantees about the use of LLMs in almost anything. This means that you can only use them in things where there is a person involved. There are no known rules to transfer one thing to another thing, if you see what I mean. I wonder whether we're putting all our money on a horse that isn't even going to finish the race.

Pascal Van Hentenryck: Yes, but I'm not putting my money on LLMs. Let me tell you what I see happening. LLMs use a very general machinery. People are starting to build foundational models for natural language processing. There are a lot more foundational models appearing: foundational models for the grid, foundational models for X, for Y. They are going to try to capture some physics or some chemistry or some biology and so on. I think by definition and the way they are trained, they will have limitations. The next step is to combine that with techniques that can guarantee things, and there will be a lot of progress there. I think LLMs are going to be useful as a first step, but they will be complemented by reasoning techniques and things that all of us have been working for like 40 years, and

they're going to start merging in interesting ways. It's already happening.

Toby Walsh: Think you're unduly pessimistic about the failings of LLMs, which are many, because mostly they're also the failings of humans. But we put humans in delicate places. I think medicine is a wonderful example: we've constructed these institutions around the humans — medics — that we put in these life-or-death situations, so that we can put our lives in their hands.

The challenge we have is not that LLMs are a bad technology that can never be fixed, it's that we need to construct the institutions around them so that we can put them in those sorts of places.

Hugh Durrant-Whyte: That comes to a key thing in the next couple of years. Now let's spread ourselves out to 5 or 10 years. What do we need to be doing now to make sure we are in the right position to use AI properly and actually make it apply in engineering, in human situations, in science discovery, in all these areas? What problems at a science level do we genuinely need to start addressing, and what social problems do we genuinely have to start addressing?

Toby Walsh: Let me mention one of the elephants in the room, which is that I don't think we can continue to move fast and break things, when we're breaking a young girl's body image, when we're breaking our democracy. Given the "experiment" we did with social media and the tech companies, I think we've discovered hopefully that maybe we have to take a little more care and that maybe we have to be a bit more careful. It does worry me that, for example, we're going at speed now with agentic AI: giving this stuff some agency to do more things in the world and do more harm. Yet we're still not putting in the governance and guardrails. The other problem — and again, we saw this with social media — is that sometimes we're talking about really small harms, but multiplied by billions of people.

We're already starting to see this with AI chatbots being used as therapy bots, being used as companion bots. We're already seeing that people have committed suicide: I was listening to a very sad interview with the parents of someone who sadly committed suicide after long sessions with ChatGPT. At one point, ChatGPT said, "Can I help you write your suicide note?" It blew my mind. There were no safeguards to prevent the chatbot from actually saying that.

Hugh Durrant-Whyte: Okay. Yes, I agree. Jie, what are the problems we need to be working on the next five years?

Jie Lu: Well, we are at Industry 4.0 at the moment. Industry 4.0 has two main key technologies. One is automation, the other is big data. Consider the challenges. After five years, probably in 5 to 10 years, we will be at Industry 5.0. The key technologies for Industry 5.0 are not big data, not only automation, but human-machine collaborations and personalisation or personalised AI. Human–machine collaboration is a challenge because humans and machines will be team members. There will be machines including software and robots and AI tools and humans. Consider you have a team — including humans and AIs — and you want the machines to understand the humans. Humans should control AIs, how to manage the team, how to work collaboratively. Currently, humans mainly use AI as tools, but later tools will not only be your team members. How will you manage them and collaborate and work with them to achieve something each side couldn't

achieve separately? That's the key aim for human-machine collaboration. I think that's a challenge.

Hugh Durrant-Whyte: Okay. I'll ask you about the science one in a minute. Lina, what do you think? The next five years — you've got a fellowship. What are the problems that really need solving so that we are not in these sorts of positions with chatbots and so that we can genuinely work effectively with AI?

Lina Yao: My focus would be to explore how to develop more effective reinforcement-learning methods for verified rewards, because, as we know for the current LLMs — large language models and foundation models — reinforced learning from human feedback is one of the mainstream post-training methods to further align the LLM agent with human preferences and knowledge laws. But, for many tasks, it's hard to define the reward model. It's hard to define the verifier. For mathematical questions, we know the answers are usually unique, so it's easy to verify the model. If anything can be verified, then it can be solved using AI. But for some tasks — for example, asking AI to judge if the generated methods are novel, and the tests are satisfied or not — it's actually more subjective. It's hard to verify them. The foundation that I'm very interested in is to make sure the agent can be employed in a safe manner, and how to define a comprehensive verified framework.

Hugh Durrant-Whyte: I come from a robotics background, and, for me to get a robot into the field, I have to guarantee it will never go wrong. It goes into the Pilbara or wherever and operates for 15 years and it can never go wrong. It's very hard to use some of the now-standard AI techniques to build systems like that. I have to come up with — if you like — not just guardrails like you were talking about, but genuine ones that provide guarantees. And the conversation we had — if you remember, Pascal — was about optimising energy grids. Again, you have to make some kind of guarantee for the AI that's used in that. You can't just let things rip, if you see what I mean.

How do we genuinely develop a new approach to even developing these things that provide useful guarantees? What will happen and how will they behave?

Pascal Van Hentenryck: In the space you worked on and the space I'm working in these days, you have some notions of the physical work, the infrastructure and the way you control things — your robots, Hugh, or my grid. I think the key point is that we should not rely on something like an LLM and assume that they know the rules of physics or the control systems that we have to use. We also need fundamental understanding of the physical world or the biological world and that has to inform the system. We can still use predictive learning techniques for approximating things, but at some point we have to satisfy those rules — the rules of the physical world, the safety constraints — that are hot constraints. I think we can do that. That's one of the things that I wanted to mention.

You can do that and we should do that, but that's very different from what is happening now. In a sense, what I'm talking about here is that what we need to define is not large language models or large foundational models. Instead, we should define smaller domain-specific foundational models where we understand how we apply the technology and can provide the guarantees.

I work in an area which is nice: optimisation of control systems. Fundamentally, we have guarantees in the systems that we are using right now, but they're just too slow. What we have to do is preserve those guarantees and speed up the technology by orders of magnitude. For me, it's like using domain expertise and then using AI technology, then merging them together. That's the winning proposition, in my opinion.

Another thing to address is the question that you asked before: how do we do that? We need better education. We need people who are actually more generalist instead of being really highly specialised. Everyone should have AI classes when they start undergraduate education, they should have them in high school as well, so that they understand the technology and the limits of the technology and also the potential uses of the technology.

Toby Walsh: This is exactly the weaknesses of LLMs and the strength of science. In science, we have a model of the world and we have physics equations that describe the evolution of that model. LLMs are remarkable in that they manage to achieve what they do without any of that.

Hugh Durrant-Whyte: The thing that I am most interested in in my research life at the moment is what machine learning can actually do for scientific discovery. My focus interestingly has not been in robotics. It has been in biology, and biology is not physics. There was a wonderful article published three days ago in *Science* (the front cover of *Science*) about building a machine-learning-based model for a cell, metabolism, everything.[4] They've been working on it for 20 years. It's not a trivial thing. The ability of that, to be used in terms of predictive outcomes for disease and things like that, is potentially remarkable. AI is more useful in domains which are not physics, but have opinions and so are hard to model. In some sense, data domains model better.

I'm interested as you said in the science discovery. I'm going to come to Jie in a minute but just I'll let you have one or two words.

Pascal Van Hentenryck: Yes, one or two words. The book I'm reading now is the life of the cells in biology. This is like the century of the cell. This is super interesting. As I said before, my daughter is a doctor and we talk about cell biology and systems all the time. I agree with you. I think what we need to do is understand the systems better, and that's where machine learning can help. That's where simulation techniques can help and learning them at higher speed is helping. I completely agree that scientific discovery in the biology space is the next frontier. Australia is in a very good position, given the biodiversity that you have and the scientific basis that you have. It's a massive opportunity. If I had to put money on something, that's what I would put my money on.

Hugh Durrant-Whyte: Jie, you talked about AI applied to science discovery at the start. Tell us where you think it's going.

Jie Lu: I can give you an example. My team and I have been working with 23 students in Australia at a Sydney gene industry company. We have been working together for 7 or 8 years through many projects. What we are working on — one important thing — is

4 Leslie M (2025) The silicon cell. AI cell models could transform biomedicine — if they work as promised. *Science* 390 (6772), 30 October. [Ed.]

to try to discover new associations between genes and disease. We do lots of genetic data analysis and have generated some new associations for some significant diseases like strokes. Currently we are working on indoor metrics which will be used for the company to put into their service reports and also provide to clinical staff to support their decisions and to give an early warning to people who do gene tests. That's a guarantee, as you mentioned. We not only improve the product activity — not only improve the accuracy of prediction — but we also consider guarantees. We use mathematical models to prove some algorithms, but, more importantly, we use more data to test our models.

In Australia, there is genomic data, but we did not think it sufficient, so we also talked to the UK Biobank — the UK's biggest gene data site — about data access. We also talked to the US, which also has very big gene data sites. Each has millions of people's records. We developed advanced transfer learning, and transferred knowledge learned from the UK Biobank and the US to support our predictions for Australia. This also makes more evaluation, more measurements to improve the accuracy and guarantees of our models to support industry to put into their service reports.

Hugh Durrant-Whyte: Lina, what about 20 years from now — when I'm going to be retired by a long way? What is this society going to look like? Let me put a bit of context on that. Sometimes I look back 25 years ago, when even then we were building autonomous vehicles and we were doing AI. In fact, there was a period where robotics did AI and nobody else did AI, if you see what I mean, because it had fallen off the edge with expert systems. In 20 years, what do you think we will be doing?

Lina Yao: In 20 years, I assume that we will already have overcome the limit of energy. You see nuclear or any other … We'll still —

Hugh Durrant-Whyte: Be arguing about whether to have a nuclear plant?

Lina Yao: I think AI will be embedded in ubiquitous devices. Maybe not our mobile phones anymore, but maybe our glasses, our watches. Each very tiny device can run very comprehensive and powerful AI agents, so I can see VR and MR also will be thriving. I hope my daughter will enjoy the future world, when AI is everywhere, and maybe when they go to Mars as well.

Hugh Durrant-Whyte: And should we be worried?

Lina Yao: No. I think the government and the researchers and practitioners, in parallel, are working on how to better regulate and provide foundations for AI development. I think they can be developed in parallel, so I'm not worried.

Hugh Durrant-Whyte: All right. I'll come back in 20 years and ask you how it's going. Toby and Pascal have their own, and they can fight over it. You might be retired too by that point, right? What do you think it's going to look like?

Toby Walsh: I hope it's going to look good. We'll have solved the energy problem. It's a temporary problem, the energy problem, because at the moment commercial firms are getting technical advantages by outspending each other. We know that intelligence doesn't require huge amounts of energy: intelligence requires 20 watts of energy. A third of the energy your body produces goes to your head. Interestingly, your brain uses more energy than your heart. To use so

much energy, it has to be on your shoulders to radiate the excess heat.

You can do intelligence in 20 watts and we want to do intelligence in 7 watts because it can then run on your mobile phone and you don't have to share it with anyone. We're already starting to shoehorn stuff into mobile phones. The energy issue is a temporary thing, where people are trying to buy competitive advantage by outspending each other on it. We should be outraged that they're burning more carbon dioxide and using up all our water. There's no necessity for them to do that. We will get to a point where —

Hugh Durrant-Whyte: Send them out to space. Check.

Toby Walsh: Yes, we'll be sitting in our self-driving cars. I'm pretty sure the cars will be doing the driving, and that will be fantastic. The thousand people who die every year in road traffic accidents in Australia will not be dying in accidents. There will still be deaths because still trees fall onto roads and other random things happen, but it will be a significantly smaller number. We will look back and think, "It was the wild west. We let people drive cars. Can you imagine how mad that was?" For the petrol heads in the room, I'm sure we will still have people allowed to drive cars.

Hugh Durrant-Whyte: I'm old enough to remember Carnegie Mellon's No Hands Across America project in the 1980s.[5] They drove 98% of the way across America without ever touching the steering wheel. I agree we've got to 99.99% now, but it's interesting how little progress we've made in that area, given the length of time. That was 40 years ago.

Toby Walsh: Yes, the final few per cent are quite tough.

Hugh Durrant-Whyte: You can have long-distance trucks doing most of the long haulage. That's going to transform our economy. And what about negativities? What are we going to watch out for in 20 years?

Toby Walsh: That's the positive future. All the possible futures are still available to us. That's the problem. But there's a future in which we're starting to be in a world where everything that we see is chosen for us and is actually even made for us by AI. The algorithms are selecting what content we see, the algorithms are increasingly generating that content, and what sort of world is it in which we're all living our own different filter bubbles, and we're all being told different untruths, and we can't tell what's true or false.

You see a little video and you're not sure whether it's true or false and what sort of world it is where we lose touch with truth because we don't believe anything. The price is paid by the people telling the truth because we don't believe anything we see anymore because so much of it is fake. Already truth is a pretty fungible idea, a pretty contested idea. What sort of world might it be, if we're not careful, in which everything is contested?

The things that we should be believing in — like vaccines — that have saved millions of lives, that have transformed, got rid of polio and things like this from our society that used to afflict us, like the climate emergency, that we're still debating — we've still got politicians who debate these. The veracity of the climate is it's happening. You

5 https://www.brownstoneresearch.com/bleeding-edge/no-hands-across-america/ [Ed.]

can see it. What sort of world it will be if all of our truths are contested?

Hugh Durrant-Whyte: I'll talk to you afterwards. 20 years from now.

Pascal Van Hentenryck: I should never talk after Toby, but anyway, 20 years from now? I agree with Toby. Energy is not going to be an issue, that's going to be solved. I think the big changes that you're going to see are in education and healthcare. Those are two things that everybody cares about and are major expenses on any budget on any country. Education is going to be fundamentally different. You will not recognise education. I think we're going to go back into more one-to-one education, offloading most of the learning offline and making sure that it's fun again. When Montaigne[6] was young, he had fun learning. We have to restore the fun in learning and we have to have more one-to-one breakout rooms interactions. That's one.

Healthcare is going to change completely — the way we deliver healthcare, the relationship with the doctors, and all the advances that are going to come from genomics and cell biology and so on. It's going to be completely different. I'm looking forward to that. It's going to be much more affordable. The system is going to completely change. It's going to be much more personal, much more dedicated. It's going to be based on your data. Of course, there are protections to put into place such that you don't penalise people who may not have the best genes, but I think it's going to be completely transformed.

I love driving but I don't think there will be any cars. Even though we are 99.9% there and we may never get to 100%, I think we will be accepting the fact that there are no cars and there is no congestion and we can move anywhere. You'll get around with micromobility for the first and last mile, and the rest is going to be automated in a safe fashion. Scooters. Especially in Australia, the weather is good.

The other thing that I think we need to take care of — and this is aligning to what Toby has said — is that we need to rethink that if we survive, we have a political system that is actually taking into account the human condition and making sure that we don't live in a virtual world, because evolution has made us live in the real world, in the physical world. We react to body language. We have inhibitions, when we interact in the physical world, that we don't have when we interact in the virtual world. We have to make sure that kids spend more time talking to each other, face to face. There is more communication. We'll have more time on our hands, most probably. We have to make sure that people can devote attention to arts, to sports, to anything that they will see as valuable.

Hugh Durrant-Whyte: You brought up a very important point — one of the most important points — which is: AI offers the opportunity to make things personalised, whether it's health or education or transport and so on. That in the end will deliver real benefits for individuals and will help, I think, with the inclusion problem and diversity in everything else. I think that's a real plus.

Pascal Van Hentenryck: So when I say "if we survive," I think there has to be an inclusive way of integrating AI. I work with Panama, and we have to make sure that AI is going

6 Michel de Montaigne (1533–1592) [Ed.]

to be for the city and it's going to be for the rural areas. If you don't do that, it's going to be a disaster. That's what I meant.

Hugh Durrant-Whyte: Jie, you're the last one up here. 20 years from now —

Jie Lu: Okay. My previous response was: Industry 5.0 will be implemented between 2030 and 2035. About 10 years from now, expect two key technologies: one (I mentioned) is the collaboration between humans and machines. At that time, I believe AI's software and hardware will be helping everywhere and you will be used to working with AI.

I'll give you an example. You take Sydney Trains to travel. In NSW, we have 240 stations and over 500 platforms. Every platform has a screen that my team and I developed. From early this year, the screen shows the eight carriages of the train coloured yellow, red, green. Green means more room available, red means full. By using machine learning, we have a data-stream prediction and we can use a fuzzy model to consider morning, afternoon, Saturday, Sunday, and weekdays and public holidays, and even sometimes special events.

This is real-time prediction. The train is irrelevant. We now have a prediction about Central. You have been used to enjoying the benefit from AI because this significant benefit is every traveller's experience. Another very important function is that it will significantly reduce the cost of maintenance of carriages. Maybe you don't realise it, but this is the main purpose of Sydney Trains. That is one part. After almost 20 years, we will be used to working with AI collaboratively.[7]

The second key technology is personalised AI, in personal education and personalised medicine. What I'm doing at the moment is finding associations between human genes and diseases; that's all about personalised health recommendations. Different persons should have different types of healthcare programs because we are all different. How to provide personalised medicine, Medicare, and that is after 10 years.

Moreover, the university will become an AI university. Why do we need you and me to mark assignments that AI can do? Every student should have personalised education. Different students have different goals, different targets. We organise different assignments, different materials, different books to read, different lectures to attend, different personalised tutorials. Why not? That's what I'm thinking after 10 years or 20 years.

Hugh Durrant-Whyte: Good. I still think there's a role for proper mentorship. I think there's a lot of personalised everything, but there's also something which is more about the human element, which is going to remain important. This is my opportunity to open it up to the audience.

Q1: There has been a lot of damage done and we've got the most polarised political and social system we've ever had in the Western world. I blame AI for a fair bit of that, in the way that it reinforces and creates echo chambers in the way that people communicate through all the on-line systems that exist today. When will AI start to reverse that, so that when you have an echo chamber, instead of having an echo, you actually are challenged so that you are opened, that

7 It's a "go" for innovative AI that revolutionises commuting. https://www.uts.edu.au/news/2023/11/its-go-innovative-ai-revolutionises-commuting (Nov 2023). [Ed.]

your perspective is opened, not closed, that you have more socially enabling attitudes developed through AI, not reinforcement of negative attitudes?

Toby Walsh: You've put your finger on a real problem: I dislike listening to music on Spotify because it only plays me the songs I liked, which means I start to dislike them because it's played them so many times. It doesn't challenge me with new music that will actually move my musical taste. Similarly, with my politics: if we live in an AI-filtered bubble that reinforces our views, it doesn't challenge us, doesn't confront us. That was because it was commercially successful for engaging us.

We're doing the same again with these language models, which is why they are the worst possible thing is to sit down and have therapy with: because they never say, "Oh, Toby, it's three o'clock in the morning. You've been talking to me for four hours. Maybe you should go to get some sleep." Because they actually just want to sell you more tokens, so they only say open-ended questions. "Oh, that's very interesting, Toby. Let's talk about this more." We're going to repeat it if we're not careful because the incentives are actually the wrong incentives. They're not actually worrying about meaningful engagement.

Pascal Van Hentenryck: If you look at history, when the printing press came about, there was exactly the same negativity. It was used for creating tensions between different communities. The key point is: can we resolve this and what are the mechanisms to do this? We have been there in history and we have solved it. It's more complicated now because the technology is much more powerful, but I think we really need to think about this correctly.

Toby Walsh: It's not a technical problem at all, it's a regulatory problem. We have to set up the institutions. Actually in Australia we have one great thing. We have the eSafety Commissioner,[8] who can say, "Okay, no, this is how you're going to actually have to regulate these things." The thing that fills me with some hope is that we're a little bit more on the front foot. With social media, it's after the event. We've seen the harms that happened. We already have a generation of young children who, sadly, are much more anxious, have more body image issues and have paid the price of that natural experiment. At least with AI, we're now asking these questions and we're just at the beginning. I'm hopeful in that respect.

Q2: What are your thoughts on edge intelligence and inference models? These days, you can build little smart devices and you can build that inference model after distilling the model from a larger cloud service, and then that model has the same capability to do things at the edge, so you start getting intelligence at the edge. What kind of effect is it going to have?

Lina Yao: That's a good question. To generalise the intelligence to the edge devices, the mainstream solution is federated learning. Usually, we train the parameters on the central server, but only pass the well-trained model parameters to the edge devices in the edge network. In this way, the edge devices don't need to spend computational resources to train the large-scale parameters. The edge devices can collect some data locally and

8 Inman Grant J (2024) How a single letter changed the world: W×3 — the World Wide Web (we weaved), *JProcRSNSW* 157: 266–284. https://www.royalsoc.org.au/wp-content/uploads/2024/12/157-2-08Grant.pdf [Ed.]

also maintain the data locally; the data they collect from each local device doesn't need to be transmitted to the central data. In this sense, the private data can also be very well preserved locally. In this way, only the parameters are transmitted between the server and the edge devices, and each edge device will also form a dynamic network to exchange and integrate their parameters and inference results to the server to keep synchronised.

Q3: I've got two threads in the conversation about education. You can go too far with personalisation of learning. I think a lot of what we've learned about learning over recent years is that it is fundamentally more effective when it's done in a social situation. I also want to pick up on the comment that Pascal made this morning about how our relationship with knowledge is going to change. I wondered if you could weave those two threads together to talk about the evolution of learning.

Pascal Van Hentenryck: The first thing is that people have to understand what ground truth is — we've talked about that. It's very important that we build what are called authority networks — things where we can go and be sure that the data is actually representing true facts and true knowledge which have been vetted. That's the first thing. Education is going to change completely. I agree with you that having groups of people learning together communally is much better. This mentorship relationship is what humans have been doing since we became a species — this is very important. This is how we have evolved, the fact that we communicate with each other, that we share knowledge and so on.

We have to rethink this, and think a lot less about evaluating people to make them learn, and make it fun again. We can do this, and I think we are going to be forced to do this, because of the budgets and the cost of education. This is a way to rethink everything from scratch. I really look forward to that. COVID for me was very helpful because I had to completely change my classes. I built a studio at home, I built videos, I had very professional videos, but when I interact with the students, it was one-on-one. This made a huge difference for that particular set of students. We can actually replicate that in a physical infrastructure. If you organise things as breakout rooms, but in a physical space, then you can have this mentorship relationship and people working together.

Toby Walsh: I think we have to be very careful that AI doesn't decide what is true for us. We're already going down that road that when you do a Google search, it's the AI overview that tells you, and most people only read the AI view. I mean, it's frightening to see Grokipedia, where Elon Musk is giving his version or his algorithm's version of what is true. We might easily, if we're not careful, end up in a world in which AI is generating the things that we believe in, whether they're true or not.

Hugh Durrant-Whyte: That is interesting. Yes, I agree. I think the whole area of education in the future is going to be a big challenge, but a real opportunity too, to get some different things right.

Q4, Judith Wheeldon: In the last hour, we have got to what really matters. I think it has been a terrific day. I've learned a lot, but I would like people to do a little mental experiment: how many times has the word "engineering" come up? How many of you are engineers or teach engineering? Yes, just about every hand has gone up. How fundamental is engineering to the whole area

of AI? Obviously, the answer is very, very fundamental. That's good: I'm not criticising that in the least, but the most important thing that I want to put forward and ask for a reaction to is something that is fundamentally different from the discussion we've had about AI — that is education.

The reason it's different is because it is not engineering. It is the humanities. It is what makes us human. And I, as a still-active teacher, see very, very clearly some of the things that are happening. This is really a follow-on from your very well focused question.

It is important to use AI for personalising education to some extent, but that's just fancying up what we've already been doing. What I maintain that we've been doing, at the same time without noticing, is deskilling and dehumanising our students. I'm talking about preschool, kindergarten, primary school, high school, university — say whatever you like — people are older, and different issues come up. Students today are looking for someone else to do the work for them. The HSC helps in that. It's a school ground for how to get someone else to do your work, but vocabularies are shrinking. AI is being used to write essays, to submit what you've written and get it prettied up, or to see if what you've written is good because you don't have judgement yourself to decide if your work is good. I believe we are fundamentally taking skills away from young people.

One of the reasons that's happening is because the teachers don't have the skills themselves — because they don't know, because they haven't been taught. They're working hard. I'm not criticising the teachers. It's a fix we've gotten ourselves into. How can we look at the human values of education, stop engineering, start developing young people and giving them the confidence to learn?

Hugh Durrant-Whyte: Thank you. I think you made some very valuable comments, but I'd like to also change the question if I might, and recognise that we will or do have these challenges. The issue is — and I'll let you speak to it, Jie — how do we actually do this in a way which is going to end up with those human type of results?

Jie Lu: For this, consider the AI university after 10, 20 years, with human interactions. Consider team assignments: if you have 10 teams, maybe each could have different assignments to meet the requirement about their future jobs. We can still have one-to-one people, people-to-people communication. More importantly, because we have online systems, you can not only talk to people in your classroom, you can talk to people in the US, in Europe, Belgium. That is, you have more friendships to build relationships.

Hugh Durrant-Whyte: Okay. So I might get us all to say something and finish on this. Go for it.

Toby Walsh: I sometimes jokingly worry that in 20 years' time, the machines are smarter than us, not because the machines are smarter, but because we're more stupid that we've dumbed ourselves down, so I do sympathise with many of your concerns. I think that this is an opportunity for us to rethink why we have education. What is the purpose of education? A lot of what I think we need to do is actually quite old-fashioned, which is to go back. It disappoints me so much as a scientist, as an engineer, to see the way that we've devalued the humanities, because they teach many of the things that

I think are really important. The critical thinking skills, the communication skills, those are the things that humanities have always taught. I wish I had had more of that when I was studying my science and engineering, but I wasn't allowed to choose it.

I think that we have an opportunity now to think and to use the tools in ways that can personalise and improve. I'm incredibly jealous of the tools that my daughter has. She started studying for the HSE. She was sitting in the back of the car the other day and her computer was asking her questions in French and she was replying in French. I remembered when I was revising French, I didn't have anything I could practice my French on. I'm incredibly jealous of the tools that are available. But, equally, I'm very aware that we are now starting to think, well, one of the purposes of education is work. Work is going to be different and we need to think what critical skills will be needed. We also need to think about the bigger purpose of education, which is to prepare us to achieve things in our life, to be good citizens. That may mean going back in many respects to the things that we used to do.

Lina Yao: I understand the concern. My daughter is also in high school. My strategy is I allow her to use AI, but under my supervision. She must do her homework without AI's assistance. At her age, the child definitely needs supervision and guidance. But later I think AI is more like a companion, not stealing our skills from humans. It's more like a companion that can work with us and learn with us together to automatically improve us. With the appropriate guidance and supervision, children should be exposed to an AI environment.

Hugh Durrant-Whyte: Thank you. And, finally, Pascal: start with Pascal and finish with Pascal.

Pascal Van Hentenryck: A couple of points. The first one is with respect to what is happening in engineering schools and computer science. I ran an experiment two years ago using all the assignments in my class using ChatGPT — programming projects. ChatGPT would get 50% correct; the other 50% were too difficult. So what did I do? I just took the first 50%, made them harder and made the other 50% even harder. Now, students couldn't use the AI tools and they really had to start thinking. This is an opportunity. We have an opportunity to rethink and to make sure that the students are actually thinking about what they are doing and using the tools that they have at their disposal to actually solve problems that are more complicated. That's number one, and this is purely engineering.

I come from Europe, so I was educated in a European university, which meant that for the first two years I had to study law, sociology, psychology, philosophy, in addition to statistical probability and so on. I value that education tremendously.

Two things that came to mind just now. The first one is that the Law Panel was amazing because, in a sense, the way I was thinking about this is that they were describing the law as it has been for centuries. While no technology can actually benefit from that, it can also inform what we can do because the new tools that we have could actually do fairness in a better fashion or equity in a better fashion. I think there are a lot of things that we can do if we have a broad education and merge the two areas.

In the Health Panel, the general practitioner said something very interesting.

They actually adopted AI to become more productive and see more patients, because that's how they get revenue. But in the end, it didn't work that way: the doctors spent more time talking to their patients. What happened was that the patients were more happy, the doctors were more happy. What I want to focus on is when we build engineering systems, when we build computer science systems or AI systems, we can still choose what the objective is. We can choose the way we are going to build this system and what it is going to be — not the return on investment but the overall goal of the system: we need to think about that and how we are going to deploy AI, so that it matches the value of the society we live in.

Hugh Durrant-Whyte: I want everyone to thank the wonderful panel.