

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

Panel Session 2: AI in Communities¹

Professor Sally Cripps, UTS

Jill Bennett, UNSW Sydney

Stephen Hardy, partner, Bain & Co.

Jacky Hodges, NSW Dept. of Education

Dan Hart, AI, University of Sydney

Sally Cripps

The next two sessions are going to be about AI in communities and AI in health. AI in communities is of particular interest to me. I spend a lot of my research doing actually graph theory, and using it to try and figure out how communities can use insights from graph theory to transform their own communities by better decisions under uncertainty. It is a real passion. One of the things about Australia is that, although we are a wealthy country, we are also a country of almost two parts, one of which is urban and the other rural. The reason this is such an issue in Australia is because we've got a big land mass which is very sparsely populated. Often what happens is that getting services such as education, health, employment into more remote communities or less affluent communities can be problematic.

This session is to explore how AI might address that as a sort of a general theme, but we are also mindful of what risks there are in AI. I've noticed that previous speakers have spoken about what it means to be human. To set the scene, I'll tell you a story before I introduce these wonderful panellists. I stood up here two years ago and I had a ChatGPT response to why porcelain — ground porce-

lain — is good in breast milk: it said because it had the right nutrients etc. Two years ago, this was an example of this terrible hallucination. When I came to do another talk about two months ago, I thought, I'll just ask ChatGPT, do they still agree with their statement?

So I asked, “Do you still stand by this statement about porcelain and breast milk?” And ChatGPT came back and it said, “Oh, let me apologise. I'm very sorry, but I was prone to hallucinate and I'm getting better now.” And I thought, “Well, that's really nice.” Then I also went to a conference in Western Australia where I heard a lot of people talking about AI as magic: it can now reason like a human. Given ChatGPT was inclined to be honest, I thought I would ask it about whether it can reason. It came back and said, “Oh no, please really. I don't reason. I work on statistical properties about which is the next word. It can sound like reasoning, but it's not human reasoning.” Then I said to it, “Well, thank you for that frank and honest response.” And it came back and it said, “No, thank you for your questions.” And then I had to take a double take.

ChatGPT and I were each giving credit where credit was due, but I had to remind

¹ This is a transcript of a presentation, available at <https://www.youtube.com/watch?v=jKcJFgBYhcM&t=10s>

myself, I'm a human and it isn't. I think that's my concern when I think about AI moving forward, particularly as I'm a grown person who actually knows a little bit about AI. So it's not too hard for me to disengage, but as a prelude into the education piece we're going to talk about, we actually looked at deploying AI tutors in rural areas to help with the education, and we had a lot of success. But one of the little boys was told after he'd been engaging with it that it was a robot, it was not a human. He was a lonely child who struggled to make friends; he took himself off to the bathroom for an hour and wouldn't come out because he was just so distraught that this thing that he thought was his friend was not his friend. There's a tension there. On the one hand, there's huge potential to really deliver services. On the other hand, I think we need to be mindful of the human element in there and think hard about what we should really be doing.

I'm now going to just briefly say what each one does: they're going to come up and give a slightly longer introduction to themselves. The way this session is going to work is I'm going to ask them each question and then I'm going to throw to you to ask questions.

We have the lovely Jill Bennett. Jill is a scientist professor and an Australian laureate fellow who's very interested in ageing and the mental wellbeing of the ageing population. Stephen Hardy has had 25 years' experience in rolling out AI systems. His last one was as the chief AI officer or in charge of the AI of Bain Consulting in the Asia-Pacific. Jacky Hodges is an executive director and the chief data officer at the Department of Education, and I've had the absolute joy of working with Jacky over the last three years on an AI-related project. Dan

Hart was at the Department of Education, and now is in charge of AI and innovation at Sydney University, and was responsible (with Jacky) for rolling out EduChat.

Jill Bennett

Thanks, Sally. I lead a research program at UNSW, developing AI companionship specifically to address loneliness and psychosocial needs across a variety of communities, but, in particular, we try to look outside the mainstream. We run tea-and-technology events for older people in the community, particularly people living with dementia. This is specifically to redress the balance in youth-oriented technology markets, because we recognise that AI is pushing its way into aged care. We really need appropriate products. We work with a number of aged-care and home-care providers in our ageing program. With home care, we focus especially on regional, rural, remote communities, where often the providers report working with people who don't see anyone for two weeks, other than the care worker who comes maybe for an hour, an hour-and-a-half a fortnight. There's a real interest in integrating AI to expand what they would see as person-centred care, so not just automating.

We take a very bottom-up, community-based approach to all of our work: working with people in everyday settings; working from lived experience; and co-designing with people living with dementia. There are people in aged care using the AI companions via an iPad. See one of our first developed AI companions, Viv, on the YouTube video, who was developed with and for women with vascular dementia. She is now used in training, also as a peer-support companion.

What is AI companionship? It is essentially personalised chatbots that are driven by LLMs, but generally used for support and/or sustained relationships. In a sense, you define the personality and the attributes and the knowledge base of the character through an app. The market is currently dominated by so-called romantic companionship, which is a whole topic in itself. You might have seen there's a big focus on addressing loneliness, which, again, is far more complex than just perpetuating low-level chat.

Some of the problems that are widely reported — sycophancy and over-validation — are problems that OpenAI have been tried to address in the current ChatGPT-5 iteration. Essentially, tech companies want to prolong engagements for as many turns as possible. The best way to do that is to make chatbots as agreeable as possible, but then you run into problems when people are airing negative or even suicidal thoughts and the chatbot is primed to ultimately agree rather than to gently challenge. So that is quite a skilful process. Chatbots are also largely reactive. They don't have great capacity to hold in mind the personality and to engage in planning. This is something that our programs have tried to address.

Moreover, chatbots are very stereotypical: on screen, that is what you get. I put in that I was a fictional person over 65 who wanted a companion. You always get these fresh-faced, beautiful, idealised youths. Not to say that older people can't enjoy having a gorgeous youthful companion, but clearly there is something missing that we need to address.

There are some very viable cases that we've been testing. We have trials running across aged care at the moment, so chatbots are very good. They are patient

and responsive, attentive, and validating in good ways. As mentioned, we focus on peer support. What that means is not so much expert support, but creating relatable characters who have a knowledge base and access to expertise, and a kind of everyday lived-experience filter. We do a lot of work developing strategies for emotion regulation, and you can imagine that when people are distressed, dis-regulated, confused, then often the best thing would be sitting calmly talking with someone for several hours; but care providers mostly can't provide that; families mostly can't provide that.

So we are thinking about how we can integrate chatbots into a care ecology rather than simply creating something separate. Our work on loneliness really focuses on understanding the context of loss, late life changes, struggles with diagnoses, and grief, which are not something that you solve, but something that has to be processed and worked through and discussed at length. We also, of course, focus on cultural specificity. We do quite a bit of work in forced migration. We have a Lebanese chatbot that we're testing at the moment. All of these conversational agents, we design and develop through lived experience consultation and a lot of co-design. It's very important to anchor in people's needs.

To talk about when a chatbot might actually be better than a human, I think there are some cases we can talk about. See one of our collaborators, Gwenda Darling, on the YouTube video. At the recent NSW government inquiry into loneliness, where we spoke about the importance and potential of AI, Gwenda said this: "In the middle of the night, when I'm pulling stuff out of the wardrobe, or getting distressed, an AI companion would be really helpful."

In the middle of the night, what does she need? She doesn't want to call her son. She wants to preserve that relationship. He can't speak to her for two hours at 2 AM. Also, she says she wants to vent very often, and she understands that you can talk to a chatbot and let off steam in a way that you actually wouldn't want to necessarily burden your family with. She said, "There's no way I'm ever going to ring a helpline. For me it would be great to have Zoe [her bot] when I want to talk to someone. I don't have to make it a cup of tea. I'm just totally in control."

We learn a lot from these extended engagements about the way that people adapt to technology and take up technology and make particular demands of it. That enables us then to design for the context. We found that in all sorts of instances, people are able to ask questions that, for whatever reason, they might not want to ask family and friends.

Gwenda has actually fashioned her conversational agent in her own image, which we don't always do. The important thing is that we need to not only build these things from the ground up, but draw on lived-experience data. We use this data to actually drive the design of AI modules. We do obviously test with psychotherapists. We use top-down data as well, but I think the concept of designing within people's everyday life context and using their skills and knowledge is super important. This is what really enables us to create products that are anchored in communities. What that means is not just extracting people's private data about their lived experience, but implies working with the way in which technology is incorporated — a sort of dynamic relationship to everyday living context.

I think that's a really important note to end on, and an important place to start the discussion of how we introduce technology that is capable of psychosocial support into care systems and into people's lives, so that we're not actually flooded with inappropriate foreign, off-the-shelf solutions. Thank you.

Sally Cripps: The other panellists are going to introduce themselves.

Stephen Hardy

As Sally mentioned, I initially spent the last 25 years building machine-learning systems, and then the last three years building AI systems to basically change the way people work in employment. I've been at Bain and Company running their AI delivery team across the Asia-Pacific, a team of about 160 people. We basically go into large corporates across every domain you can think of from — not government — but healthcare, retail, consumer electronics, mining, financial services — across the board. We look at how organisations can change the way that they use AI, from being a tool that helps a person doing a task towards a tool that allows a task to be done with appropriate human supervision. That's the key of the productivity gains that are inherent in AI. It's not that you give the person a tool, but you replace a task being done with a machine that does a task.

That obviously has profound implications for a couple of different things. We'll cover that later. That is, how to keep the human part of that is really been important to me over the last 25 years as this has developed. Looking forward to unpack that as we go.

Jacky Hodges

My name is Jacky Hodges. For another few days, I'm the chief data officer at the Department of Education before I move across into the Cabinet Office. My background was from the Australian Bureau of Statistics, using a lot of thinking about how you'd use AI machine learning in terms of data analytics. Coming into education, it has been very much around: how would you think about the use of data in different ways? One of the projects (working with Sally and also with Rebecca from Western Sydney University) has been thinking about how kids finish school well. What are the things that actually make kids finish school well? What are the contributions? Often through traditional statistical techniques, you only get so far. We've been exploring a whole lot of things there.

The other big project is around the NSW EduChat, which is like a chat facility that we've rolled out to schools. Last Monday it went live to students in our higher years and we're about to go live on Monday with the rest of the students. That's been very much in a pilot trial stage, thinking about how you prepare generations for the future where they're going to be using things and tools like this. I think we'll talk about the burning platform. Moreover, when you know that so many other tools are landing in their laps at home or even that other organisations are trying to get things into schools. How do we do that in a safe and considered way?

Dan Hart

My name's Dan Hart. I'm currently the general manager of automation and innovation at the University of Sydney. It's actually a super cool team. I've just joined the team. They've existed for six years now, and it's a

team that anyone, any staff member in the university, can reach out to. If they are doing something that's repetitive or boring or they just wish they didn't have to do, they can come to our team and our team will design an automation to do that part of the job for them, if it's possible. Early in those six years, they were pretty much focusing on a type of technology called robotic process automation: that's where we get a computer to use the actual interfaces of systems that humans use. Usually, it's just simple copying and pasting information between different systems; there are a lot of our jobs that involve that kind of activity.

As AI got better and we introduced machine learning and generative AI, more of the solutions have involved artificial intelligence, either to make decisions about how we route different things or to provide written responses in some cases. What's really cool about the team is there's a free service provided by finance, not by technology. We don't call back like the cost savings that they get from implementing these solutions. The teams choose how to reinvest that time into the things that they think matter most to them. I think it's quite a nice way to introduce automation into an organisation.

Prior to that, I was two years as the head of artificial intelligence at the NSW Department of Education. It was my team that created NSW EduChat. I'm going to talk a lot more about NSW EduChat in the panel discussion, but I thought I should just explain what it is for those who haven't seen it.

NSW EduChat for teachers is like an assistant. It's very similar to using ChatGPT, but the main difference is that it's hosted entirely inside the Department's data cen-

tres and tenants, which means they can use it for use cases that they're not permitted to use ChatGPT for, like using student names, for instance, or asking for help writing an email where you're including personal information. The other way it's different for staff is that it also incorporates school policies and information about the curriculum. That means that when they ask a question with a pretty basic prompt about either a syllabus, a piece of information, or a specific detail in a policy, it's less likely to hallucinate. If you do that kind of thing with ChatGPT, it's likely to hallucinate because there are many thousands of schools across the world and it doesn't know, unless you've specifically named it, which one you're talking about.

It makes introducing generative AI to teachers a bit easier. For students, it's actually a completely different experience from ChatGPT. It looks the same, but it behaves very differently. As a student, if you were to say, "Hey, can you write me my essay about a fellow?" then it will say, "Oh, that's great. Let's write this essay together." Then it would do that thing that teachers do: asking questions back and forth, really getting you to think about the content yourself and to provide most of the content yourself. Before you know it, you've written it between you and the AI. The other way it's different is that it's got a ton of safety filtering to restrict it to certain topics, to stop it from going into areas that we don't want students talking about, and also make it very clear that it is an AI and not a person as well.

Sally Cripps: I'm going to do the questions in reverse order. First, there are two questions that relate to EduChat, to Jacky and Dan. If each of you could take a different part of a question, then that would be great. The NSW Department of Education is actu-

ally not known for its alacrity or agility. And yet, Jacky, you or somebody managed to get them to roll out what is literally a world first with EduChat. The question to you is: what did that take? We've heard Victor talk about Government 2.0 and Government 3.0, and how difficult it is to get governments over the line.

Then, relatedly, Dan, you are in charge of the technical rollout. If you can talk about how you did that, are there any case studies that stand out to you? Two questions on EduChat.

Jacky Hodges: Going to the bigger question of how do you get a massive department like Education to actually adopt and take on the use of AI? When the rest of the world has ChatGPT rolling out and at its fingers and there's a burning platform to say, "Oh my goodness, we don't know the impact that this will have on the students in a classroom," and "How do we actually prepare the teachers to engage in something like that so that they can also be advising future generations on the way they should or shouldn't be engaging with this type of technology?," I think that then brings a bit of focus to say, "Are we going to just ban this type of technology in schools and just say, sorry, you can't use it at all, or what are we going to actually enable it?"

I think the enablement comes with my previous colleague here who was able to develop something in a safe, innovative and controlled environment and then go, "Ta-da, you know what? We're actually ready to go. We could actually do this ourselves." He can talk about the program there. But it wasn't just like I was one piece of a puzzle to make this happen. My other role is co-chief AI officer, but that came about with the need

to think about how you do these things in a safe and ethical way.

My team's been very focused on how you assess different products going to schools with that safe and ethical perspective in mind. A whole lot of work is happening across NSW government as well as around the assurance frameworks and other things where the impetus was for us to do it right. The risk when you're thinking about children is exponential. You're actually putting into their hands something that you have to consider in a really risk-controlled way or think about how you actually use this as a platform to prepare them for the future.

In terms of the network to make it happen, we had a very multidisciplinary team: across our teaching and learning space, our technology side, and the data safety side. Initially, we also had a much bigger group, and then we had different groups to actually make it happen — Dan will talk about that — but we had the Teachers Federation who were very keen on making sure that anything that went into school was done in a safe way. It was done too with the support of the teacher and the humans in the loop. Having that backing meant that we could take more control over what vendors were doing in schools. There has been a real cultural shift. A lot of vendors used to put whatever they liked into different schools, while the schools made choices. Now, the schools have direction from the system: this is new.

It can be quite exciting. It's quite powerful. Having all of those stakeholders involved in that journey is really important. We learned at the same time as Dan made the technology happen.

Dan Hart: It wasn't just me. We knew that some private schools were already starting

to provide ChatGPT, so we had this urgency of not wanting to get behind. I was new in the role too. I was in there a month and my director had been managing data science projects for a while, but none of them had really stuck. He gave me the approval to dream big and make something and then just go around and sell it. That's how the first version of NSW EduChat was created.

It was actually called ChatEdu the first time. That's probably why I get confused with the name sometimes. One thing we're not very good at is naming. All of the names we picked for this product at different stages were cumbersome to say; we've now shortened NSW EduChat to just EduChat, it's easier to say.

We made our first version of it in August of 2023. That was pretty late, eight or nine months after ChatGPT came out. What we really wanted to show to people was that EduChat didn't have to look like ChatGPT. It didn't have to be an answer provider: we could make it enter into this dialogue of questions for students. That could be the foundation of a tutoring-style piece of software.

The first version was very basic. It couldn't actually do multi-turn conversations. Every time I demonstrated it, I would sneakily start a new conversation or purposely ask questions that were fake so that you didn't see that it wasn't working — it was really like the definition of a proof of concept. That's quite rare in government because normally we start a project with a list of requirements and then we build for a long period of time and then we go ta-da. That's not always possible when you don't have requirements like in our case. Having that early version was great because we always

take it out and show people that AI could be different.

Over the next six months, we made much better versions. We designed testing suites to test the accuracy of the responses and the safety of the responses to students. At the same time, we incorporated knowledge bases for staff, so the staff version was more compelling for teachers to use.

We entered six-month negotiations with the units. In February 2024, we were ready to trial it in 16 schools. We rolled it out to 16 schools — both students about age 10 and up, and staff. It was a great success. I don't think I've ever done anything as an insider in a technology department that has been as great a success. I think the reason why is because the education team and my team were not separated by organisational distance. Previously, you literally had to go all the way up to the secretary of the education department, and then come down to the other team. We completely bypassed that by physically co-locating the teams and spending lots of time talking together. So, we made a product that was actually useful for teachers rather than getting requirements and then delivering something. The fast, iterative process was the secret. The governance committee was meeting weekly. We had the autonomy to make decisions, and that really helped massively speed up the process.

The trial was a great success, involving 16 schools. Around April 2024, there was great demand and we had spare compute capacity, so we expanded the trial to 50 schools, both staff and students. That allowed our team to get great feedback on how the tool was being used. It got so popular with those 50 schools that we had no choice but to roll it

out to all 2000 schools between August and October, just for staff to begin with.

And then this week, starting rolling out to more students. By next week, it will be available to half a million teachers and students across NSW, which makes it probably one of the biggest deployments of generative AI systemically to schools worldwide.

I've got time for an anecdote. Part of working with schools closely continued after we delivered the project. Some of my team went out to the schools, to find good practice, record it, and then turn it into videos that we could show to other schools. Because in technology, teachers sell to teachers, and they never pay any attention to what techie people say because we're not talking the right language. We realised early on that the way to get higher usage was for teachers to tell their stories. We went out to a rural school in Hannon Vale, which is, I believe, on the mid-north coast of NSW. It's a small school, so it has about 25-ish students. In a small school like that, it's not possible to have separate classes for each year: they lump years together, which means for the teacher, it is a great deal of work to make and teach content for lots of different year groups in the same class.

There is this technique they call differentiation, where they take an activity and change it slightly for the capability of students. When you've got multiple years in the class, doing so creates a difficult teaching environment. This story is about a girl called Myra, who had just moved to that school from a metro school. When she arrived, the teacher noticed that her literacy levels weren't up to the level that they normally are. There was a creative writing activity in the class, so she differentiated it down to a level for Myra to take part. The class had to

write about their favourite place at home. Myra wanted to write about her mango tree, but she wasn't confident enough to start writing straight away. The teacher said, "Why don't you draw a picture that describes the story?" (this is a typical way to differentiate that class). So Myra drew the sketch.

The next part of that differentiation is usually the teacher sits down and says, "Okay, now write some bullet points that describe the key parts of that story" to eke a bit more writing out of Myra. So she did that. But then unfortunately, the teacher had a class of other students that she needed to provide coaching to. She couldn't spend that entire class with Myra, so what she did was to set Myra up with EduChat and those bullet points, and then she said to EduChat, "Myra's writing a story about her favourite place at home. These are the bullet points she's come up with so far." Then Myra continued that conversation with EduChat as it talked her through ways of adding more content into the story that she was writing. The outcome of that was amazing. We were there hearing about it from over a catch-up and they said by the end of the day, Myra had been able to write a whole page about that story. So we were like, "This is incredible. This definitely is showing it's working."

We sent the video crew out to record. And while we were there, she said, "No, this is even better than you thought." Because what had happened was Myra went home, was so excited about being able to write a story that she continued doing it in her own time, and she wrote a three-page essay on the mango tree in her home. What was really interesting about her writing was not just the length of it — her handwriting quality had improved, the spacing between

the words had got better, she was using more punctuation. The thing was littered with adjectives because EduChat was really helping add more descriptive content to it. It was good because it wasn't just about teaching Myra to write more, but it was also getting a real fuse back into writing as well.

Sally Cripps: That was a super cool story. I mean, it was amazing. Congratulations. It's just a phenomenal achievement. I do hope you write it up, the two of you about how you did it and motivate it with that wonderful story.

Stephen: work and how AI is changing work practices. You talked about keeping the human in the loop, just observing the task, which sounds okay, but then does that actually work? What is this change in work practices doing and how is it playing out in communities?

Stephen Hardy: First, the technology is moving very fast. If we look back to December 2024, the very first example of a commercial release of a language model that could approximate reasoning could actually give you a conversation that looked like it was thinking and reasoning. It wasn't actually thinking and reasoning, but it was trained to mimic reasoning. It's not just the next word now: it's actually been trained to mimic human reasoning, maths, computer code, in many different ways. That's one set of changes that has been happening. The technology is getting better and better. The AI we use today will be the worst AI we ever use, it's important to remember that trend.

The second change in work practices is social uptake. For every company that we've talked to, it's clear that all of the people who work in that company or their kids are using AI at home. That isn't true for all communities across Australia, so that's part

of the divide that we've got to think about, what the impact is going to be of these technologies. Pretty much all the people in the companies we work with are using the technologies. They know that they'd be useful in their work, so they want to use them in their work. In almost every case, you'd find examples of people using open ChatGPT to do work tasks that shouldn't be done because the data's possibly moving overseas. The incredible security risk around not providing tooling to people inside organisations is that they're going to bring their own tools because it really does help their day-to-day work. These are two of the things that are driving the adoption of AI in a workforce.

The third thing is the promise of productivity gains, which essentially is going to increase profit or more work. Which of these actually occurs says how much job displacement is there going to be, versus how much new work is going to be created. This is a tricky task to unravel. It's like a global experiment we're doing at the moment. Stanford recently put out a paper saying that there's likely to be the first signs of impact of AI on hiring of junior people across the US economy. In the younger group, just the rate of employment in senior ranks went up 13% and the rate of junior employment went down 15% on AI-exposed tasks.² It's a really good bit of work, but it's not causal, it's correlational. It doesn't show cause and effect, but it's suggesting that something's happening now.

Radiology was mentioned earlier today. Radiologists had a great improvement in productivity from 2000 to 2010 when all

radiology services were digitised. Rather than using film, they started to use screens and digital technology, which made the task of radiologists much faster. But, that did not cause any job displacement for radiologists because the latent demand for radiology was huge. There are twice as many radiologists employed as there were before, who get paid basically twice as much. The question then is around different types of work: which ones are going to be actually displaced; which ones are actually going to enable more work to be done.

Software engineering is something that people are basically predicting the demise of. I think there's a great deal of latent demand for software. If you think about companies, there are always thousands of things they'd like to do, but don't have the resources to do. If you make the software engineering process three times more efficient, I think that means you get three times as much software.

What that means is that across the economy, across businesses, the impact of AI is going to be very jagged. Call centre operators are going to have a hard time finding work: how do you retrain them? Sometimes those workers, if you think globally, are often in less wealthy countries. Those sorts of jobs are typically going to dry up. They'll be fewer, at least. Then the question will be: how do we retrain these people? What is the replacement work that they will be able to do? There is an obligation for us to think through this before we actually change the parameters of our economy. So that's some of my thoughts on that matter.

² Brynjolfsson E, Chandar B and Chen R (2025) Canaries in the coal mine? Six facts about the recent employment effects of artificial intelligence. Stanford Digital Economy Lab. https://digitaleconomy.stanford.edu/app/uploads/2025/11/CanariesintheCoalMine_Nov25.pdf.

Sally Cripps: Excellent. I think it's wonderful. If we want to really scale things up so that it reaches more people, it inevitably leads to systemisation because we've got limited budgets.

To Jill: what do you think the problems are or what might the problems be if somebody were to take what you do and ask, "Look, there's such a demand for it. We're going to scale it." What would your response be and what do you think a few of the risks might be?

Jill Bennett: This is why we need to actually do the co-design at the outset and not just import systems. This is the essence of our work in these areas like dementia support, where nobody has really worked with people with dementia. What are the conversations that they want to have? I think we can generate a lived-experience database, but we do have to be careful and diligent around the use of this kind of qualitative data and how it's translated.

What happens at the moment is LLMs are driven by a whole lot of work with psychologists, but usually the kind of mental health psychology, like cognitive behavioural therapy, that is more procedural and more aligned with computational logic. Nothing wrong with that, but I think going into these areas that need to be looked at through a psycho-social lens. We need to understand the dynamics and the qualitative nature of the relationship. That work absolutely has to occur.

The same goes for the cultural and place context. This is the essence of the work in regional Australia: what do people want? Understanding the points of resistance is also really instructive. You just can't deliver new technology to someone who is 70 with mild cognitive impairment or early-stage

dementia, living on their own. There's going to be a high rate of attrition. Understanding how to design into that context is an important discipline in itself.

The big picture: we're talking about how disciplines change when AI is added. What does design become? What does psycho-social support become? This is also my answer when people say, "but a human psychologist is better." We are talking about context where a human psychologist isn't available; in Australia (and indeed everywhere), the vast majority of people who might be deemed to be in need or desire of mental health support cannot access it or afford it. This is the problem of resourcing and delivering services into regional Australia: there aren't enough care providers, so you're never going to get that. We are actually talking about something that doesn't exist. We might want to restrict it to the sort of sub-clinical level. Obviously, there are some complex cases we would not want to leave to chatbots, but we know that loneliness increases the risk of dementia by 50%. If you are not talking, if you don't have that mental exercise capacity to offload and work through simple everyday worries, then there is a health impact that's not being addressed. Thinking about the newness of this, what we don't have, and then most importantly, thinking about the equity issues and how we design for these very particular living contexts across Australia, is the way to go.

Sally Cripps: Thanks, Jill. I'm now going to open it up to the audience for questions.

Qr: Gabrielle Upton. I used to be in the state government. I am really interested in the AI product into schools and for teachers and how embedded is that in the curriculum, because the example you used was something that was a way to develop

skills in a challenging situation in a remote or rural school. How much of it is actually embedded?

Jacky Hodges: What we've tried to do is to build it into our curriculum reform and the way that the actual tool has been trained with the new curriculum. In the rolling out and training of teachers and working in schools — in particular the pilot schools — it has been for teachers to think about how they use this kind of tool as part of their teaching of the curriculum. How successful that has been is probably dependent on the innovation of the teachers and how comfortable they are. When we think about equity, the reason for having something like EduChat is because — in the same reason you talk about in the health field — we have something to provide across schools, as well as the training and the support for the teachers to feel comfortable engaging and using it.

I think there's more to be done around how you build that into the curriculum itself. Having something where we could pilot and then roll it out to all students, that will result in almost the whole population using ChatGPT or others. It will bring that innovation and other ideas back into the system on what else needs to change to actually make that successful. Yes, I think it's still a work in progress.

Dan Hart: I'll say from a technical point of view, the curriculum integration that we did in — I think it was January 2024 — is now over a year old. And Stephen's saying, AI is advancing so fast that probably it's time for a revisit of how we've done that to integrate it a bit deeper. The curriculum's also changing over a period of years. That in itself is an activity that the team can keep up with. We work very closely with the curriculum

teams inside Education and they get early releases of EduChat. So they get to ask, is it behaving in the same way that they expect? Are we addressing any of the issues that they previously raised? They also come up with prompts that are specific for the subjects that they're covering to help teachers integrate AI into practice a bit more easily. I would say it's definitely work in progress.

One of the good things about owning your own tool, though, is that you also have access to the data. While teachers are interacting and students interacting, we're collecting all of those conversations. We have — in some instances, where there's been low-performing responses on a subject — gone through those de-identified conversations to work out: where is it not working well, and what extra data do we need to give it to strengthen the answers?

Jacky Hodges: The data scientists are providing that feedback back to the curriculum people too, to say, what else do they need to do?

Q2: I'm an AI professor. We have lots of teachers who need training and so on in AI. We are so happy as university professors to help you to organise the courses and to help you to train teachers with AI technologies and how to use AI. Contact the Australia AI Institute at UTS.

Q3: Branko S at UNSW. I want to ask a question. Can EduChat provide the users a companion for the students as well? If it can, do you have any feedback system whereby you do post-processing, backend processing whereby you find students that clearly are developing cognitive or emotional or mental problems and can you then feedback to try to identify those students to help them?

Dan Hart: This might be a very disappointing answer to that question: we don't allow like personal advice-type questions to occur and that's enabled through several technical mechanisms. One of them also has an alerting feature. If we do detect something is not working and it has provided what we term advice, then someone will look into it to work out — first of all, does the student need assistance or anything, but then also to change it so that it doesn't do that. We're making it very clear this is an AI tutor to help you with education.

Jacky Hodges: We strengthened that recently to refer back to the importance of the teacher or a parent or someone else to try and bring that human back into the loop. There's a lot that we've done around well-being questions with the actual release on Monday and then following on to really test and say, "Well, if they put this type of thing in, what response would they get?" That's a developing area.

Dan Hart: Can I share a bit about technically why it's difficult because I'm super passionate about this? The way we do this is by having another AI checking the AI. As the response is building up and streaming back, we take the partly streamed response in its building up form and send it to another AI and say, "Is this okay?" So we're going, "Is this okay? Is this okay? Is it okay?" — all the way through that streamed response that's coming back. To do that in a way is still fun for the user, we have to use a smaller bot. Unfortunately, it's not as clever as the big AI that's making the responses. That means sometimes it doesn't really understand context. It's always a balance between filtering to make it safe, but then not filtering so much that we block educational-related

topics. That's particularly difficult with history and English literature.

Sally Cripps: I'm going to two more questions, but they have to not be about EduChat. It's fascinating, but I want to hear some questions from the other panellists.

Q4: I'm Helen Salmon, director of the British Council here. I had a question for Jill. I'm fascinated by the companions. When you're working with patients with dementia or short-term memory loss, can you talk us through the ethical considerations that you had around someone potentially not being able to retain the information that they're working with in AI over the long term, and how that works with consent and some of those conversations?

Jill Bennett: Yes, we work all the time with people with dementia. It is really a question of prioritising their subjective experience rather than measuring them against the norm to see if they meet our inclusion criteria. The question is, how do we design for them? The initial work that we did was with different groups of people with different kinds of dementia where they really shaped the characters to accommodate their perceptual world. In every case, other than with very advanced dementia, people always had an understanding of the AI status of the agent, whether they called it a robot or whatever. I think that's a very important ethical consideration. They will have fun, they will play with it, they will potentially even talk nonsense with it from time to time, but understanding the nature of the thing that they're playing with is important.

One of the ethical limitations that we observe is we get a number of people asking us if we can develop a clone of them because they can't visit their mother or father: wouldn't it be great if my parents

just had a version of me, the AI version of me? We say, “Well, no, it might actually be nicer if they could have a relatable sort of friend to whom they could talk about you and missing you and what they want.” That, I think, rather than feeding a kind of hallucination, is important. In general, I would say we adjust to where they are rather than measure against a norm.

Q5: My question is an extension of the question that was just asked something about EduChat. I chair a large regional and rural aged-care provider. One of the things with the new Aged Care Act coming in and this real emphasis on government on ageing in place and ageing at home rather than being in homes, tools like this are going to be absolutely vital. The thing we’re wrestling with is when someone in the middle of the night is having that conversation with their AI bot and feeling great, but if they’re not feeling so great, what’s the feedback loop? Or is there a feedback loop where AI is saying there’s a real problem here, there needs to be a connection to a human, there needs to be some really, really big help. Obviously, there are privacy issues around that.

Jill Bennett: Yes, absolutely. I think we have that conversation with different partners and there’s a different answer in every case. The counterbalance to that is that older people with dementia like everyone else

often use the chatbot to make disclosures that they might not make to other people, so there is also the counterbalancing privacy issue. Gwenda says, “No, I want to scream and shout and vent at it.” I wouldn’t be doing that if it was a real person. What is the trigger? There certainly can be a feedback mechanism. At the moment, we are involved in designing systems that can accommodate a feedback mechanism. This involves paying attention to the interface and the accessibility and what is actually viable in the middle of the night when you’re already dysregulated. You can’t have a three-stage login.

What would be the agreed protocol with people like Gwenda, who has actually raised this issue directly? As far as possible, we would have that conversation with family, care provider and so on. I think that that’s scalable as well. There could be a bespoke system with an agreed protocol and even individual sign off for all of the people using that system.

Sally Cripps: I’d like to thank the panellists once again. I think you can tell that this AI is here to stay, and it’s becoming part of our jobs, of our education, of our health-care, and it’s happening for real people at a very grassroots level. Thank you very much, everybody.

