- Hello and welcome everybody. First of all I'd like to acknowledge the traditional custodians of the land on which we meet, the Gadigal People of the Eora Nation, and pay respects to their elders, past present and future. I'd also like to welcome all of you, and thank you for coming, and particularly thank you for our speakers coming and joining us today, and also we have people who are streaming online, so welcome to you too. I'm Zach Thomas, I work for DFSI in the ICT digital policy area.

- And I'm Katie Ford, I'm from Data61's government team. This is the second of these events that we have been running Zach, the last one was in November.

- November, yeah.

- On the topic of data sharing. Hello Kate, thank you for coming. So this event will be focused on AI, and the purpose of these sort of events, is to really try to de-demystify. So technology is such a funny area, we often have a lot of conversations with people, and there can be a lot of blank stares. So this is really trying drill down into some of the concepts, so what exactly is AI, why are people getting so excited. But also from a government perspective, does AI have to be a black box? And I think we'll hear some very definitive ideas around that today. How can we make AI explainable, how can we make it reviewable? And from a government perspective in particular, how do we deal with the question of ethics, we're gonna have a whole presentation on ethics today. What is fairness? I think IBM has classified 200 different types of bias in human decision-making, when it comes to AI machine learning in particular, those biases become explicit, so how do we deal with those? We have a stellar line-up today, I'm very excited, our challenge today between Zach and I, is to try to keep everything on time. Please save up your questions, because we will have a good chunk for Q&A to the panel. And I'll hand over to you Stewart, bit of an intro on the Data61 agreement with the government.

- Yeah, first of all a bit of housekeeping. The bathrooms are in mezzanine area beyond the muffins. The emergency exits to the sides, and to the back of the building, of the room. As you'll have seen there is refreshments, and there should still be some of that left when we leave. And also this event is being live streamed and recorded, so if you have a problem with that, please approach Felix at the back there, wave Felix, thank you. So if you've got a problem please approach Felix. And the evacuation procedure is the normal sounds of the whoop whoop, and if the main fire alarm goes off, do not use the lifts. So with no further ado, let's talk about the agreement quickly. So we would like to mention that we have a whole of government agreement with Data61, it's a quick and agile way to work with them and access Australia's leading data specialists. There are some pamphlets at the door about the agreement if you are interested, or you can approach me or Katie, or Melissa or Jacob are also here.

- And we'll put up our contact details at the end if you want to reach out to us.

- Yeah. So first up we have Pia Andrews, Pia is the Executive Director for Digital Policy and Innovation for DFSI, she's leading our team of policy design data development and engagement experts to drive the digital transformation of the New South Wales government. And she brought with her to us a wealth of knowledge working with technology in private public and community sectors, with a focus on transforming governments and enabling innovation through digital public infrastructure, so please welcome Pia, thank you.

- I'm afraid I'm a mad gesticulator, so I always need to have my hands free. First of all, thank you all so much for coming, I think that AI, and generally a whole bunch of different emerging technologies are very exciting, but I think conversely they are also a bit terrifying for a lot of people, so how many people here are terrified by AI? A few? How many people are excited? Interesting, okay, cool. So we've got a lot converted. But it is worth knowing that and realizing that a lot of these new technologies give us real opportunities to actually transform life how we know it, to actually transform how we live, to transform society, and so then it's our choice as to whether we end up having our values wrapped around technology, or technology wrapped around our values. Really deciding about what the future is for us, for our societies. Whether we are designing for everyone, designing for ourselves, designing for past paradigms, or designing for future paradigms, that's our choice. We went through an industrial revolution where a whole bunch of jobs got fundamentally changed, and we invented a new paradigm, we invented trading in money and data and information, and created a whole bunch of new jobs, we can do the same again. I think the real challenge in this space is trust and dignity, and I'll come to that in a moment. But artificial intelligence is probably one of the most over-hyped things at the moment that you can find, people refer it to everything, same as block chain, but at the same time, the genuine opportunities are extraordinary, and the challenge I set to all of you, before we get into the exciting aspects of today, is to think about as you listen to everything, and as you leave this auditorium, how you're gonna use your new knowledge, your existing privilege, your existing work, to do good and to help shape that better future, that better more inclusive future for everyone, and how we can shift paradigms. I see personally, not as an AI specialist, but as a great observer and participant in technology revolutions, I see three waves of AI happening at the moment. Mostly its reactive, how do I take a current job and automate it, you see a chat bots as the poster child for AI, which I find really interesting, because it's nowhere near where it gets interesting. A lot of people are starting to think about how can you do proactive AI, applying machine learning and get an AI to engage. My favourite example of that actually is in Taiwan, where they used parent AIs to spend out myriad mini AIs to go out and engage people in sentiment. What do you feel about a change, how do you feel about Uber coming into the market, and then we're gonna use that sentiment to distil down into actually inform in real time, publicly develop, collaboratively develop with community the legislation to allow Uber into the marketplace, in a way that also maintains the values of that society. That teeny is just an extraordinary use of AI to enable better democracy. And then of course we get to the whole point that I think gets missed a lot when companies and governments look at how they can use AI. Is the personal use of AI, all of these tools gets into the hands of the majority of everybody at some point. And the notion that how we make ourselves, in government, in companies, in NGOs and in research organizations, open to someone else's personally tethered AI, not provided by us, to engage with us, that becomes a very interesting question. How do we make our rules, our content, our business systems, our transactional systems, our information, our knowledge, all available programmatically so that personal AI can assist a person who may be very vulnerable in some cases to make decisions about their lives, and understand how they interact with a system to be empowered to operate and have agency within the system. It's very exciting for me personal AIs where it's all at. Anyway, so some of the critical factors to think about today and moving forward, and we are gonna be talking about some of it, is absolutely the ethics of this. I don't know how many people saw the ATO legal decision a couple of weeks ago? Go and have a look. There was a very big, big decision that we'll have to fix at some point, but right now is I think going to the High Court, which basically said that a letter generated by the ATO was not considered legitimate or binding, sorry? Oh really? Ooh okay, so we've got a problem. I'm not gonna say this right, maybe this gentleman can. But basically it said that because a letter was generated by a computer it wasn't considered authoritative, of course that creates a problem for automated decision-making, and for many of the decisions already made in the economy. But there is a real mix here of, is a person more accountable than a machine, and how do you make machine decisions accountable. For me, this is where we get into traceability. How do we ensure that the point of the decision you have, explainable decisions you have, accountable decisions you have, immutably recorded decisions, that a person has access to all the decisions about their own life and the data about them, so that they have a bit of agency in that. How do you ensure, and this goes into legislation and regulation as code, some of the work that we are driving, and one of the leading jurisdictions in the world now, looking at how we could actually get to and API.Legislation.NewSouthWales.gov.au. Not just translating after the fact, but drafting, getting test driven regulation. Test driven legislation. Anyway, personal control over data is critical to this, modular architecture is critical to this, transparency and traceability and personal visibility is critical to this, if it's gonna maintain trust and dignity into the future. And of course we also got to look at how we use AI to enable a post-scarcity paradise. We can't just keep speeding up the past, and expect to get a different outcome. So there's a lot to hear, there's a lot to learn from, I'm really excited about the day. I'm really delighted that we can record these sessions, so that they help inform public narrative and debate about these opportunities. But also about the role that we can each play in shaping these opportunities. So a quick thank you very much first of all to my team, to Zach and Melissa in particular, thank you so much for putting this together, and to our support from the rest of the team. Obviously a big thank you to Data61 for putting this event together, and for being such a great research part of the general story nationally. And a big thank you also to our media folk, and Felix and the team, thank you so much for recording these things, we really appreciate and you do an amazing job. Research is critical, how we engage in research is critical, and how we fund research as research I think is a big question we also need to address, and that's gonna be a fun one to tackle as well. So welcome, thank you for coming, have fun, and then with your great new power, make sure you use it responsibly, thank you.

- Thanks so much thanks Pia for making the time today, we know you've been flat out in your role, and we are really excited to be working with you and your team at DFSI. It's my great pleasure to introduce Adrian Turner, CEO of Data61. So Adrian came back from Silicon Valley I think in 2015, after 20 years over there as a tech entrepreneur, he came back with a lot of ideas, a lot of passion, and he's really excited about driving change across the entire Australian economy. So thanks for coming today Adrian, and I'll pass it over to you for your presentation.

- Thanks. Alright thank you. So what I'm gonna focus on is the intersect of AI and the economy, and talk about some of the implications for Australia, and why this is so important for us as a country right now to get right. So for those of you not familiar with Data61, so we are the data science arm of the National Science Agency. Our mission is to solve, or help solve Australia's data driven challenges that are underpinned by deep science and technology, and we do it in partnership with the University sector and others domestically and around the world. I just want to frame when we talk about AI what we mean. It's one of these terms like innovation, that means a lot of things to different people. As a label it's been picked up and used in a lot of different ways right now. This diagram is imperfect, but it does give you a sense of the different components that we think about machine learning, natural language processing, robotics, computer vision, and I'm sure that Richard and Tiberio will talk more about some of this. The reason AI is so important to the world and to the Australian economy, is that it is a general purpose technology. So if you think about other general-purpose technologies in the past, you can think about the transistor, pretty well everything with a CPU and a power supply can trace its way back to that original invention of the transistor as a general purpose technology. Internal combustion engine the same, and is not just the invention itself, but it's the second and third order consequences on the economy. So you think about cars and how that drove urbanization and people moving out of cities, so structural change in the economy, and the ability to reshape and reframe supply chains once we had the internal combustion engine in trucks and other things. So digital innovation today accounts for 11% of the world's GDP, and its growing. And we recently did some work with Alpha Beta that looked at the Australian economy, and compared us across four core dimensions to our OECD peers, and those dimensions were digital productivity, so the application of digital technology to drive productivity gains. Multifactor productivity, so the re-engineering of business processes as a result of the digital technology. The creation of new domestic industries. And then the last one being the creation of new digital exports. And what we concluded, was that we are 34% behind our OECD peers across these four dimensions, and if we make structured steps towards closing that gap over the next decade, that represents 315 billion GDP opportunity for Australia. And the place that we are performing least right now is digital exports, we're 20% of our OECD peers, and we think we've got a role to play in resolving that together. So it's not only creating the right sort of regulatory frameworks and environment to allow that to happen, but it's also accelerating the translation and research into outcomes and impact that can then be commercialized globally, and making sure that we are funding our research sector properly to be able to focus on the longer term problems, and the deep science and tech problems that will feed into those solutions. So we identified eight high potential areas for Australia. Some of these will not be surprising. Precision health, digital agriculture, data driven urban management, a big one. Smart cities. Cyber physical security, so cybersecurity as a lot of things, but cyber physical security, the intersection of the physical world, or with the physical world, we think Australia is uniquely suited to go after. Supply chain integrity, important, because 97% of our businesses are classified as small businesses, and they're gonna need to connect into global value chains. We think about food, we want to trace our food through the value chain to be able to command a premium for our product overseas. Proactive government, so Pia touched on some of the potential there. Legal Infomatics, which Pia touched on as well, so Deloitte pegs compliance a $249 billion drag on the Australian economy, so if we can automate aspects of that, there is tremendous payback. In the US it's about a trillion. So by solving it for us we also create an export opportunity, and smart exploration and production. So this is autonomy as an example. We think at the core sits data competitiveness, and we define that as data capture, so this is sensing, this is robotics. Data management, so how do we deal with data, manage data, secure data, share data across organizational and jurisdictional boundaries. Data analysis and the different techniques that we can use for analysis, with trust at the core, transparency, explain-ability, big challenges to be solved. And then decision and action, so how do we deal with the human interface and the behavioural aspects, and making sure we get the right data to the right people at the right time. I also want to touch on this work that were doing, which is close to wrapping up, so we are being asked by the federal government to lead the development of a national AI roadmap, as well as ethics framework, and we've done that through broad consultation, right across the country with different stakeholder groups. So the rationale there, and I'll come back to this, is as an economy we are too small to not get nationally organized around AI and how we go after it, and I'll come back to that in a moment. So with a strong sovereign capability, the outcome and impact for us as a country is huge. So we talked about it in dollar terms, but we can think about it across three dimensions. First one is the economic output, and the economic impact. And that's a lot about GDP and productivity gains. And this is some work that we are doing, we spun out the group called Emerset, that's using LIDAR technology to be able to self navigate drones in GPS denied environments, so think about dropping those down into caves where people don't want to go or can't go. The second one a societal, so this example is work that we've done in the past, to put sensors on the Sydney Harbor Bridge to help with predictive management of that infrastructure and predictive maintenance. But the whole potential impact here for cities, if we think about different types of infrastructure for the sea, whether it's water, whether its energy, whether it's transportation, whether it's people movement, whether it's environmental monitoring, and being able to derive insight across those data layers becomes incredibly powerful going forward. And then the last one, and Pia touched on these cycles of AI, and that right now what were doing is trying to make things more efficient and automated with the current frame of reference. So looking at what we've done in the past, and trying to do it better and faster and cheaper. But the real opportunity is to unlock entirely new value, and what this is, is a program called Lion's Share, it was conceived by a Sydney production company, and its using machine learning and computer vision to identify animals in TV commercials and films, and actually pay them royalties for their participation, and then feed those monies back into protecting their habitat. So from a small idea, this is taking off around the world, and the group has convinced David Attenborough to front it, it was just launched at the UN not long ago, and we were involved with that. There's big brands that are now on board, like Mars and others that are coming on board, there will be a big announcement in the next couple of weeks. A great example on whether science and technology partner to Lion's Share. Doing something that was never possible before, to create entirely new value. And just in a global context as I finish up here, is the world is investing, other countries are investing enormously in getting organized here, so whether it's the UK, whether it's France, the EU, India, Singapore. Of course China and the US are investing huge amounts here, and they're doing it because it is a general purpose technology, there will be economic advantage to the countries that leap ahead here. And I think the opportunity for Australia is not just a focus on the economics. We're a country that's got an egalitarian culture, an inclusive culture, and I think that plays very well into the ethics discussions, and I think it's about capturing the economic, societal and environmental benefit, but doing it in a way with ethics at the core, and then exporting that thinking to others around the world, and in the same way that people trust our food, and trust our legal system, there's no reason that we can't be known to be trusted for our AI as well. And one anecdote, just to highlight how important this is to finish. If we think about the Internet, the Internet was conceived as an open communication system. And at the time a lot of people wanted it to be free, the unintended consequence of not thinking about there needing to be an economic model to fund the build out of the Internet and the applications and services, is that it's now evolved to be a system or network that in part surveil's us, if we think about the advertising network, and the deep socio economic profiling that they're doing of us. So this is an unintended consequence of not having the right conversations upfront, at the birth, as the Internet really moved, from DAPRA and other places into the mainstream. So we should learn from that, and be having these conversations right now. And if we get this right, Australia is so well-placed but we don't have a lot of time, the rest of the world is moving on this as well, we need to get organized nationally, we need to optimize for the country, not for any one institution, not for any one sector, this is a Team Australia all in, we need to get together and lock arms and pursue this. Thank you.

- Thank you so much Adrian. We wanted to do a little bit of a deep dive into some of the live case studies, what are the genuine questions, what are the genuine areas of challenge that researchers and companies around Australia are working on in AI. And I could think of no one better than Dr Richard Nock, who is a principal research scientist at Data61, and an adjunct Professor at ANU. Richard where are you Richard? I lost you, oh there you are. Thank you for coming in today Richard, I hand over to you.

- Thanks very much for the invitation. So to be honest, I'm actually a little bit scared, because when the audience was actually asked this question at the beginning of who is scared by artificial intelligence currently, there was a quite substantial number of people who said yes, so I hope this number is not going to increase after my speech. Ideally the excitement is going to increase, but let's say the fears is going to decrease. So what I'm going to essentially develop in this talk, is a short story of machine learning, in which I'm going to put the work and the objectives we are trying to solve, so there will be essentially four parts to my talk. I'll first talk about a very brief story of what happened at the very beginning of machine learning. It's very important to capture what has been happening at the beginning to essentially understand what's happening today. And then I will make a brief description, of essentially the kind of problems we are having today, we are having to solve. And then there will be I believe a few take-home messages from what's happening in machine learning, to keep and think about. And I will end up this talk by essentially a short forecasting of what could happen in the near future for machine learning. After all, my job is about making prediction as accurately as possible, so we'll try this one. So to start with, before essentially going backwards, let's start by a very short crash course in machine learning. So it takes two slides, there will be no equations. So the idea is to understand what is machine learning about at a very high level. So what is typically happening is that we have a machine, this machine is not necessarily a computer, it can be a robot for example. This machine receives signals, I will call this signal data. So these can be sensors, these can be human activity, this can be human input, this can be a medical record and so on. And this machine has a task, and this task is to essentially come up with a model, so this is the brain here. And this model can be very simple or extremely complicated, and the objective of this model is essentially to give insights, and so for the machine to give outputs based on what it has learned. To achieve some task, to predict something and so on. So this is the end of the course. So now let's have a quick look at essentially the way it started. So this started very long between quotes time ago. And essentially as much as artificial intelligence was initially motivated by the study of games, machine learning also started with some very nice and some very simple problems centred around games. So in artificial intelligence, essentially the first games were for example the game of checkers, and these were actually the first games for which the computer led to some solutions that were essentially beating all human players. And in machine learning we had this very simple problem of tic-tac-toe, which people actually developed, it's just one example of course, it's not the only one. But this is one very simple example of the domain with which people actually developed their first, and tried their first learning systems. In this case the input is just a set of tic-tac-toe games, and the objective of the model, of this brain, which is of course very simple in this case, is essentially to predict whether the configuration is winning for one of the players. So that's a very simple problem. And if you understand that this is indeed a very simple problem, then you will probably accept the idea that machine learning essentially started in the sterile room. Because the problem and the data was so simple that they were nothing compared to what was in fact happening in the real world. And so this idea of the sterile room which is important to keep in mind, and there is also this idea of what happened in the evolution at the upscale, which is essentially millions of years ago. An era, a geological Earth era, which was called the Precambrian, and during which you had in fact a lot, actually not a lot, but let's say a lot of very primary lifeforms, which later on evolved. And then it all started this way, just like in the sterile room, just like in this Precambrian era, you had no predators for example. And then, so 30 plus years ago, so what we had back then was essentially access to very simple clean data to summarize. The objectives were very simple, the models were very simple, the assumptions were very simple, and we had also a very small number of problems, they were all kind of similar, the models were very, very similar, they were essentially a set of a handful of different models. And this was all far from today's complexity, and then of course came the problems. Because then essentially what machine learning had to face is essentially the throwing of these techniques into the jungle of the real world. Or let's say what happened after in the Cambrian area, essentially the predators developed. So let's have an idea, let's have a look at the kind of troubles that we can meet essentially today, and from which I hope you will understand that the substantial part of machine learning today is about coping with problems that could be very simple if we didn't have additional constraints that essentially make the problem very difficult to solve. So let's have a look essentially at a subset of course of these problems, there are so many problems, makes researchers happy of course, but there's actually a very large number of problems to solve. So you have essentially three kinds of troubles that you can have in a machine learning system, depending on where it's located. Either in the input data, or on the system, or in the output. So input, the system itself is somehow the engine, so you have also the output of the system. So let's have a look first at what can happen with the input, and then I will develop very shortly what we are doing in this area. So suppose that your input consists of images, imagine that your system is actually for example an autonomous car. Or an autonomous robot. So as input you're going to receive for example, sets of road signs, and of course looking at these images It makes absolutely no doubt for the human eye that they represent very concrete signs, and nobody at least being trained for driving will make a mistake for these signs. Now this is something that can happen. So somebody, this person right here, has access to your database. This person is not going to steal your data. This person is in fact going to modify the input to your system. So either the data with which you train your system, or then the inputs of your model once it has been trained. So this person is going to modify the data, and if somebody tells you that your data has been modified, and then you have a look at your data to check whether this is indeed true, this is what you are going to see, and then again for the human eye, maybe it's gonna be very odd knowing what the background images can be, it's going to be very hard to confirm whether indeed you had some unexpected access to your data. But then when you feed this images to a system, this is what the system predicts. So the system essentially becomes completely broke, and the predictions become completely wrong. In this kind of context they can become obviously, the consequences can be extremely dangerous. So what we do, is in fact, so in Data61, with a set of partners, I've put the partners here, is we investigate this system. So what we do essentially is what you have here, so we essentially somehow build an attack on the attackers somehow, a system that is going to prevent the attacker from making substantial modifications, and we do this in the context of robotics, because it's one of the examples of applications of machine learning in the real world, that's really going to need this kind of techniques. When I say robotics, it's essentially the field in which you have a robot Who knows nothing about the task at hand, just simple rules, and then the robot essentially has to learn the way it's going to interact with the world. So this can be an autonomous system, and the autonomous system is going to be put in some conditions to learn, first he's on the way to interact with the world, and then progressively the interaction is going to be more and more complex. Imagine what could an attacker do in this kind of context, so we are doing research with some university and industrial partners, and with the Australian government as well, to imagine some countermeasures essentially to this kind of interaction by an adversary, so this is called adversarial training. Now let's have a look at another problem. So still this problem is about the input to your system. So let's say your input is a set of medical records that you keep somewhere in the database. And this time what happens is that somebody logs on your system, and they don't modify your data, they just steal your data. And this is the kind of, let's say thing, that we see every week around the world, this kind of data breach is happening absolutely everywhere every week around the world. So what can you do with that? Well there is a very, between quotes, simple way to find a workaround to this problem. And this consists, again we figure out some sort of way to prevent these kind of attacks. We just encrypt the data. So essentially we replace the data on the hard drive, for example by encrypted data. So the data as it is currently stored on the system looks like garbage. So obviously somebody coming and stealing the data, we don't reduce the chance that the data is gonna be stolen, we just guarantee that if somebody steals the data then this person is not going to be able to figure out what's inside. So it's obviously a very simple way to prevent these kinds of problems. But then the question comes, can you learn from this kind of data. Because then as I said your data looks like garbage. It turns out that this is possible, and this relies on very sophisticated training techniques, again without this constraint, the problem could be fairly easy to solve, but with the constraint that now your data is represented in your computer in the form of this kind of garbage, this is much more complicated to do, and this is something that we do in Data61, in a project called confidential computing with state and government departments. So now just to finish this list, I'll just quote some other problems that can happen, and we in the machine learning group are not necessarily focusing on these problems, but other people in Data61 actually do, so what can happen at this time on the machinery? So let's say on the models or on the algorithms. So the constraint that you can have today, are constraints that can be related to energy or hardware. Energy, because your algorithms are going to be carried out on huge data centres, they are heavy consumers of energy, and you want essentially to reduce or constrain this footprint. That actually really changes the problem, and in the same way, your model which here can be extremely complex, you may want to actually implement your model On a very simple cell phone. So how are you going to be able to transfer your model, and make sure that such a big model can be run on a cell phone Without requiring too much resources, still allowing you to call for example or use your phone as it's supposed to be. Then this is another problem on which people are actually thinking about. And to finish up with this list of problems, there are also problems related to the output of the system. So when I say the output, it can be the model, it can be an prediction, it can be an action, so it's very general again. On the output, this time you have questions, and these are not new questions of course, about fairness, about bias, about explain-ability. So the bias question is essentially now becoming extremely important, because people discover at some point, that the machine was so good at making predictions, that it was also very good at learning the bias which was in the initial data. And so comes the question of how we can correct this kind of bias, finesse this kind of abuse. Explain-ability is something we work with home affairs in Australia, and this is in this case how you can present to a user who is not actually an expert in machine learning, how can you present this users the results of your algorithm in such a way that he's going to understand what's in the output of the system, and he's going to be able to take as rapidly as possible the best decisions based on this output. So this is another problem we focus on. And if I can summarize all this in a final slide for the trouble part, then essentially you can have troubles everywhere on your system, so this constraint as I was saying, as the beginning that make machine learning problems much more difficult, you can have them everywhere. I have put privacy in the input, but you may also want your model to be private, not just your input data. You may want your model if it is stolen to be totally non-understandable by an external user. It can be on the output as well, you may want your model to actually send via the Internet a prediction, and if somebody takes this prediction out of the Internet, you don't want this person to be able to see what's in that the prediction, so privacy in fact can be everywhere, and this is essentially the same story for all the other constraints, and of course you have additional constraints that can essentially touch the system in every step. So legal constraints, trust constraints, ethical, transparency and so on. There is in fact no limit for this list, and it's actually very great, because machine learning is now touching on the actual problems of the real world with the actual constraints. So the take-home message of this speech is somehow, and this is something to keep in mind, that a long time ago, and a long time ago is just 30 years ago. A long time ago problems in data were very simple, and machine learning was easy indeed. Easy within code, because essentially a problem could be hard, but we knew that it was hard, and there was no purpose in trying to solve this problem. Otherwise it was very simple to get a model, and for the tic-tac-toe game it was super easy to get a good model and make good predictions. And this is unfortunately not any more the case. And the thing to keep in mind, is that the outcome when you add these constraints, can be very, very different depending on the constraints themselves. So depending on these constraints, you have a first case in which you have credible solutions that exist, but there is some work which is needed to put them at work, this is in fact the case in privacy, so these algorithms I briefly talked about, when you want to hide your data and then learn from data you cannot see, these algorithms are heavy and time consuming, resource consuming, so there is a good deal of optimization that still needs to be optimized further to get something that can be run at scale. But people have no doubt that within the next few years this will come at scale indeed, and this will be used in the real world for some very large scale problems. The second case, you have existing proposals, they are substandard, everybody agrees that they are substandard, but we have hope that they can be solved, and this is the case for the adversarial problems that I was mentioning. In this case many of the protection techniques that we have today in fact are substandard, they really restrict the way the machine is going to learn, and this is not something which is satisfying, but we believe that there are better solutions ahead. A third category is a set of problems or constraints for which we don't know solutions yet, but some attempts exist. And this is the case in the legal area. For example when Europe actually put in place the RPGD, businesses had to find solutions for this kind of problem, and we are very far from a solution that would make the internet work in the same way as it was working before the RPGD, as it is working now after the RPGD, but there is hope. And attempts exist of course, because businesses they have to carry still there business, so there is research in the area as well, and attempts really exist, and something is going to happen. And the last category, and this is important to keep in mind, sometimes also for some constraint, we know it's proven that there will be no solution, and this is the case for fairness. Some fairness. You have lots of models and constraints related to fairness. When you put too many constraints, essentially you're not going to have a solution to your problem, and we know that already. So there will be some need for some compromise in this kind of area. What is the kind of fairness you are looking for? Is it possible to indeed get a solution from this set of constraints? And so the take-home message is to essentially figure out the availability of machine learning for a given application. There is first a need to properly formulate the related constraints, and one corollary is in fact for the machine learners to properly solve an application, they in fact need domain experts to properly formulate the constraints. So that's in fact something which is very, very important. And now before ending this talk, let's have a quick, very quick look at what could happen in the way forward. And for this, I'm going to put until its end, the analogy with remember this Precambrian analogy. So this slide is actually based on a paper about the Precambrian transition which recently appeared in Nature. So I'm going to give you a brief snapshot of the actual story, as it is indeed believed and recorded now, and this is kind of the science fiction part, at least for a subset of it, but you will easily get where I want to go. So in the actual story, what happened when the transition was made to the Cambrian, is essentially that oxygen concentration rose globally, and the equivalent for machine learning, is this is what we have been seeing over the past few years, is that data collection explodes globally in this case, so you have more and more data, and very rapidly, very rapid increases in the amount of data. So what happened for the Cambrian, is that you have a substantial set of evolutionary innovations that were not existing before. Vision, legs and so on that are developed. And in the field of machine learning, what we are seeing is that you have a set of what is called combinatorial innovation, that's output to the stage to solve these problems, they can be mathematical, they can be related to the hardware, you have lots of different ways to solve these problems. Then what happened in the Cambrian, is that life essentially conquered the third dimension of space, the first animals were essentially not moving, they were staying in their 2D space, and they were happy in quotes about that, and in machine learning, essentially what's happening is that machine learning conquers the ubiquitous dimension of text. Machine learning is getting everywhere essentially. And then what happened in the Cambrian transition is that you have of course a very large number of creatures that actually get to appear on the Earth, and in particular prey and predators, these were not existing before. And of course the food chain complexifies. And what happens in machine learning is that you have a boost in of course the amount of dollars that's being put in the system, workforce as well, conferences are literally taken over. One of the main conferences last year in machine learning was sold out in 11 minutes. So in 11 minutes there was not any more ticket to be sold to the audience. And now let's have a look at the consequences of all that. The Cambrian transition in fact resulted in the establishment of most metazoan phyla, which are the groups of animals. From a very small number of ancestors, these 2D animals somehow. And in machine learning what we are seeing is that today, and this is happening today, we have the establishment of many major classes of problems, including this constraint in particular, from a set of very few ancestors, which I talked about at the beginning of the speech, but what's most interesting is what happens next. So in the case of the Cambrian transition, what's happening is that the descendants of these animals, actually they rule the planet today. And if we made the same kind of prediction for machine learning, and actually it's starting to appear, is that the ones, or the businesses, the organizations in the position to solve these problems will in fact rule the data and the tech planet, and this is what you can really observe already with the likes of Google and Amazon and Facebook and so on, each of them is very strong in an area where the others are not essentially, so they are figuring out their niche, and essentially they have a ruling position in here, so this is somehow happening as well in this space. And that's the end of my talk, so I hope you're not scared any more about machine learning, but maybe more excited, thank you very much.

- Thank you very much. Next I'd like to introduce Nathan Frick, who's from Revenue New South Wales, he's going to talk to us about his experience in exploring the potential for using AI to drive better outcomes in New South Wales government, thank you.

- I've been sitting here deciding whether to go with the microphone or hide behind the lectern, but I think I'll follow suit and go with the microphone, so bear with me. Good morning, my name is Nathan Frick, and I'm from the project management office at Revenue New South Wales, we are part of the larger DFSI family. And I'd like to thank our hosts, Pia and her team, and Data61 for the invite to come and speak to you today about my experiences in AI over the last little while of revenue. There's a bit of a story, as I said I come from a project management office, I'm not a technology person, I don't work in IT. But about six months ago 12 of my colleagues from revenue, were nominated to go on a leadership acceleration course, a fantastic course, it was very valuable. But of course you don't get offered these things without a price, and the price was we got divided up into teams and we had to deliver some projects, and build some business cases for projects to solve some particular pain points that the organization was experiencing. So the 12 people were split up into teams of three, and each of us were given a little bit of a project to go away and solve. And three of the four projects that the teams were allocated were very specific pain point problems that the business needed to solve pretty urgently, they were very tightly constrained. My group, the three of us were given a very broad research task, to go away and talk about how artificial intelligence could be used to improve government services for customers. So it was very much a case of which of these things is not like the other. We had this broad, very open ended research topic to cover off. So the three of us turn to each other and looked at each other, and one of my colleagues leaned over and said, what's AI? So we were coming off a pretty low base, the three of us had pretty diverse experience, I'm a huge geek, so I have read a lot about AI over the last few years and the rise of AI, so I had some background information. One of my other colleagues said, well I've heard a lot about what's been said around AI in the news, Elon Musk, and statements about the end of the world and stuff like that, that's about the limit of my knowledge. And the third one said, I've seen all the Terminator movies, and that's about it as far as my knowledge of artificial intelligence goes. So the first thing that we decided we had to do, was step away and decide for ourselves what artificial intelligence was. And basically we came up with a model that started with your classic computing automation. You program the computer, it does what the program says. We are all familiar with that, our society has been run by computers for the last 40 or 50 years. But on top of that, the new stuff is this machine learning, where the system itself derives its own rules set, and that was quite interesting to step into, and to have a look at that. And then above and beyond that I suppose, is where you actually provide the software with a sense of autonomy and agency, where that self derived rulesets are actually engaged to take action then. And so you're getting into the space of driverless vehicles and things like that. So it's sort of a building pyramid, and as laymen that's how we defined for ourselves what artificial intelligence was. A pretty simplistic model I admit. So having come up with our definition, we thought we should have a look at some applications, and we looked at some external case studies in government in Australia and elsewhere that had employed artificial intelligence techniques in specific cases, and we looked at what worked really well, and some of the things that didn't work so well, and what the drivers behind some of that was. And we took away a number of really big lessons about that. First off where you are employing software with agency to take action, you need to be very, very careful about how far you provide that software its sense of autonomy. Even the largest technology companies like Microsoft have developed technologies that have had to be pulled because they haven't worked as intended, so you have to be very cognizant of unintended consequences. And you probably don't want to be too overly ambitious, at least in the public sector in terms of rolling out AI. So we decided that we would take an agile approach, one step at a time, and try to build on what we had done, rather than going all in in the first instance. So we had done our research, and then we decided that we should look more broadly across our agency to see what had already been implemented in terms of AI. And surprisingly we looked across the business, and there was quite a bit of activity already going on. Analytics area were using machine learning and AI to step into predictive analytics, and they are doing some great work around trying to proactively characterize based on past data, classifications for identifying vulnerable people in a proactive sense, so that we can get some diversionary stuff around fines that people who are vulnerable and probably can't pay those fines can be diverted into other programs. And also interestingly, we've got a whole project around intelligent robotics, and robotic process automation. And given that those streams were well underway, we decided that we wouldn't look too closely at those, we would look at other innovative things that we might be able to do using AI. So we did quite a bit of collaboration around the business, got lots of ideas. We took those away and tried to synthesize them and look for synergies between them, so that we could take that agile approach, start small and build. And the one thing that surprised us a little bit was the sponsor of our project, our Executive Director for strategy and transformation was very keen that we didn't just go and research, but that we went away and piloted, and actually built and tested and experimented and did some stuff. So we carried out a few little pilots, very small scale, very small, in order to inform the roadmap that we were gonna propose back to the business. So I'm mainly going to concentrate on those small couple of pilots that we undertook. The first one was around partnering with an external consultancy who had developed a thing that is close to my heart in the project management office, which was an automated survey tool basically in a nutshell, that helped develop a leading indicator of project success. Most of the indicators that we use around projects in the project management office in government are generally lag indicators, not predictive lead indicators, and by the time you see that something is going off track, it's already off track and needs to be recovered, we were really looking for something that was a bit more predictive. And so we partnered with a consultancy who provided us a little bit of software called Meeting Quality, and one of the functions of this automated tool is its got built in sentiment analysis. So in order to get a leading indicator of project success, some of the free text that is provided as part of the response to this mini survey tool is then fed through IBM's Watson artificial intelligence engine, and returns results on the sentiment underlying those comments. And then that folds in with some other data that the survey tool collects and provides a predictive indicator on whether a project is on track to succeed or not. So we looked at that sentiment analysis, which is only a small part of the overall program, and we thought that's a really interesting application of AI, and I know Pia mentioned sentiment analysis earlier. And so we thought how else could we use that technology in an innovative way. So every year we all fill out the PMES engagement survey as part of the public sector, and those results are tabulated, and again there is a chance to provide some free text commentary in those surveys. So we thought what would happen if we took away that commentary, and ran it through the same IBM sentiment analysis, what could this tell us about how we are making our people feel, because in the old customer service paradigm, people don't remember what you do to them, so much as how you make them feel. So interestingly we got some fantastic results back from the first time we ran that through, and that has now helped us shape and inform our response to the last engagement survey. So it gives us another lens to look at that data through, we had a nice neat dataset that we could pull through, and it gave us an interesting lens that we didn't have previously. And we plan to do that going forward, with both our interim pulse surveys, and with our ongoing annual service. We mentioned virtual assistants and web chat before. This was another thing we were looking at, we had undertaken experiments in virtual assistants internally before, to help our people access some quite complicated technical knowledge in our knowledge database. But ultimately the idea was to, to stand up a public facing customer serving virtual assistant. Now again, when things are customer facing people get a bit hesitant and risk adverse, so we thought if we start internally and develop an internal chat bot for our own staff, that would be a good step in learning how to do that. And then we thought if we open up a web chat channel in our existing contact centres, we have some quite large call centres dealing with debt recovery and fines and so on and so forth, then we would be able to use AI to layer a couple of functions on there that we don't have today, for instance we could if we use web chat, have an artificial intelligence driven web translator laid over the top, so we could get real-time translation services over web channel. And we ran a quick proof of concept, there was an existing chat function built into our existing call centre systems, so we switched that on, plugged it into Microsoft Translator, and proved that you could very easily do real-time translation across that chat. This was the original internal virtual assistant trial that we did internally, and it was very well received by our people internally, that was not done by us, it was done back in 2017, but we thought that we could use those findings to build the business case for continued work in this space. So we were looking for synergy, so as I said, we started off with the idea of an internal virtual assistant, we could then deploy the customer chat channel which would give us that text data coming backwards and forwards. And the beauty of that is we could also layer potentially the sentiment analysis in real time over that, so that we could open this channel, and we could have a significant measure of how we were making our customers feel through our interaction on that channel. And then we could layer in the translation services, and then ultimately after thorough testing, we could deploy a customer facing virtual assistant, that had the sentiment analysis layered over the top, and had the translation services built in. So that would save a significant amount of money and operating costs on our current engagement with translation services, and it would give a new and potentially quite beneficial channel for customers to engage with us. In the meantime, we absolutely suggest that we should continue the robotic automation stuff that we've already got underway, we should continue to support the machine learning and predictive analytics work that our analytics team were doing, and we should keep our eyes open for opportunities to use AI where it comes baked in to applications that we use. Because increasingly I think this is going to be the way things go, particularly in cyber security applications and other corporate functions like that. The platforms, particularly when we are moving more and more to software as a service type models, AI is going to be built in as a layer to a lot of these things that we buy off the shelf, so how can we as an organization, can we make sure that we are set up to take advantage of that, and I think the key learning there is, we need to focus on data, and we need to make sure that we as a business understand our data, are collecting the right data, and managing and owning that data. So in order to be in a position to be able to take advantage of AI when it comes baked in, we really need to probably get a better handle as an organization on our data, clean, quality, well understood. The last couple of slides I have here are just on how we proposed to stand up some agile delivery of that roadmap, but I won't go into that today. So I thought I'd leave some time for questions, because I don't think I'm sitting on the panel later.

- [Katie] You can sit on the panel if you want.

- [Nathan] Thanks. Before I, does anyone have any questions?

- [Katie] I think we might put it all in the final Q&A. But I'd love you to sit on the panel.

- [Nathan] Okay, no problems at all.

- Nathan, thank you Nathan. Nathan was a bit of a last-minute confirmation, but we are so happy you could come and share your experience, I think everyone in the room would have learnt something. Now I recognize that a lot of this is very dense, so I was going to suggest that if people want to stand up, turnaround, introduce yourself to somebody behind you, ideally, not next to you, I know you know each other, find somebody you don't know, introduce yourself, I'll give you about 45 seconds or so, and then we're gonna go into a fabulous new presentation coming up, and we'll have the panel, so remember to collect your questions for the panel. So 45 seconds. Alright, I'm gonna bring everyone back. Sorry, it was just a brief relief. I knew it would be hard to come back once I let it open. Okay, thank you everybody. It is now my great pleasure to introduce Dr Tiberio Caetano, who is the chief scientist at the Gradient Institute, so the Gradient Institute is a new not-for-profit that is focused on ethically aware AI, they are located right here in this building, and actually have a range of partners, including Data61, Sydney University, and IAJ as well. So Tiberio is here, thank you for coming, and thank you for taking us through some of the ethical considerations that governments should be thinking about. Thanks Tiberio.

- Thank you very much Katie. Hello, good morning everyone, it's very great to be here to talk to you all today, so thanks for coming, and thanks Data61, thanks New South Wales government for making this an interesting morning for all of us, hopefully right. So far for me it's been cool to be here. So I am really lucky here, because there are many things I could be talking about Now that I don't need to talk at all, because it's already been well covered before. So I am going to talk about what can potentially go wrong with AI, so we've seen before that there are reasons for optimism, for excitement, but there may be genuine legitimate reasons for concern as well. So we're going to talk about that. Not about that. But we are going to talk about other issues. I'll get started.

- [Katie] Thank you.

- I'll get started. This is actually a failure not of AI, but it's a failure of some form of system. And what we are going to talk about today is failures of systems, because what just happened is a failure of the system that involves people, involves machines. And this can happen and does happen all the time. In this case it was not so serious, why it wasn't that serious? Because actually we saw it happening. We saw it happening and we were able to fix it just on time, by the way this was not staged. And the problem lies when things goes wrong and you don't see them going wrong. Just like the example that Adrian gave very early today. In the early days when I was an undergraduate in physics, we were all very happy at the University that we could use the Internet for free. No one bothered talking to economists to figure out that there is no such a thing as anything that's free, there's no free lunch. And today we live in the world that we are living in now. Indeed, we need now to start asking very deep questions about what should we be doing with this technology. How we should build it, how we should not build it. The decisions we need to make now as Pia said before, will be consequential in the future, and it's up to us to decide how to build this AI. It's not up to anything else. Remember AI is a technology, it's built by us, it doesn't have an agency of its own as I will describe soon. So here is the promise of AI, this is why I spent my days for the past 20 years thinking about this, it's because AI is an extremely powerful technology to solve problems, problems that basically destroy our lives, problems like cancer one day will be solved by AI. Problems like poverty one day will be solved by AI. All these massively huge challenges, they just can't be solved by us without a superior form of intelligence, and that's the promise, we need to get there. So this is the reality of AI, so this was the promise, but this is the reality. The reality is that whenever you create a solution for something that solution creates new problems. Who is familiar with that? It turns out that when your solution for something is elaborate, is sophisticated, guess what happens? If you're not really careful the problems that come out of it will also be complex, sophisticated, sometimes very subtle, and this is what's happening here. Now look at this, who is familiar with this case? Quite a few people by now. So this is basically Google auto image tagging brilliantly successfully detecting airplanes, skyscrapers, bikes, graduation ceremonies and cars, and then mislabelling two black people by gorillas. So it actually did some things extremely well, it detected the airplanes and the cars, and I work in computer vision for several years I know how hard that is, it's extremely hard. But guess what, that solution, that AI that was so clever, went and made that mis-classification, which is very ethically sensitive and charged. So there's a problem here, now the question is, this is a problem we see, just like the mishap at the beginning of this presentation with the low battery at 10% or whatever it was, we can see that, but the question is what can't we see? So Richard did a brilliant job at summarizing machine learning into a slide, so I'll do something even more egregious, and summarize the whole AI into a single word, and the word is called delegation. Software runs the world, and has been running the world as Nathan said, for over 40 years. Now it turns out, if you don't know that software runs the world, basically you've been sleeping for a very long time, software runs the world. Now the fantastic thing about AI, and the reason it's so powerful, is because software traditionally was written by people, and you may still think that software is written by people, you are wrong, well you're right, but you are not entirely right. Software is written by people, but most software that exists today is not written by people, it's written by another piece of software. So machine learning, one way to see it, is basically a technology that writes software. Based on what you tell it to do, and based on data. We are going to see that in a second. Human delegates decisions to an AI by providing a goal, a purpose, a problem to solve, and providing data. So instead of a programmer actually giving every single detail of how to solve a problem, you tell it, I want to drive from here to Canberra, and the data that I provide you, well here's the GPS, and here's all the cameras in my car, and if I have a self driving car, the car goes and gets the camera. There was no single programmer that actually wrote every single instruction, well if there jumps one kangaroo in front of you you go left, if there are two kangaroos one from left and one from right, you go straight, if there are three kids and one kid is 165 height, no, there's no such thing, it's impossible for any programmer to encode all of the possible things that can happen on your way from Sydney to Canberra. It's impossible, the only way you can do that is if you have a machine to write billions of lines of code that actually contemplates every single of those possible options. And that's what machine learning is doing. Now what the AI then goes and learns how to drive that car from here to Canberra, which actions it needs to do, based on the goal, taking me to Canberra, for example as fast as possible. And based on all the data it's collecting, all the camera footage around it, and everything. Now we are going to see that like with human decisions, poor delegation can lead to poor results, and this is the case for poor AI, AI can lead to bad consequences. Look at this case here, let's say you are a grocery retailer, and you want to run personalized specials, and send every week some personalized specials for your clients, your clients are registered in the loyalty program, so the retailer knows what you've been buying and when. So the decision, the data is who buys what, which is available to the company, the instruction, the goal that you provide is maximize profit, and the decision is which items I should include in my weekly letter for personalize for you. What could possibly go wrong here? Any ideas? Yes?

- [Man] You could send a teenage girl

- Yeah that's one example, you can come up with a bunch of others. Let me give you one particular one. Let's say that you have been living in the inner city suburbs, you are an affluent person, you're a vegan and you buy fresh produce all the time, what do you think is gonna turn up in your email next week? Chips and soft drinks? No? Now let's say that that's not the case, that you come from a disadvantaged community in the Sydney Western suburbs and you're not vegan, and last week you have bought chips and soft drinks and many other highly processed foods, what do you think will turn up next week in your email? More of the same. That's what the machine is learning, it's not like you were telling the machine you should do that, the machine is just learning that's what maximizes profit, after all that's the goal you have just set, maximize profit, so to maximize sales you've got historical data, and it's perfectly solving the problem that you asked the machine to solve. Okay?

Right, another example, recruiting. What could possibly go wrong if you have an algorithm that screening CVs to select which people should go for an interview? You've got hundreds and hundreds of CVs, you cannot interview hundreds and hundreds of people, you run an algorithm, you select, you need to automatically triage which CVs are worthy of people going to the interview. So that data are just paper, just PDF files of CVs, so what could possibly go wrong?

- [Man] People lie on their CVs.

- [Tiberio] People lie on their CVs. What else?

- [Woman] Middle-class white males.

- [Tiberio] Middle-class white males.

- [Man] If you're training it, it will replicate what you've got.

- [Tiberio] You will replicate what you've got.

- [Man] You just automating your existing bias.

- Alright so this is all there, and give a very simple example, let's say that the role is for a software engineer, and you are basically trying to detect who in your pool will be potentially good software engineers. Now what's the supply out there, what's the pipeline for software engineers? Well, 85% males and 15% females or something similar to that. What does that mean if you're learning from the data, it means that 15% of your training set is for women, and 85% of your training set is for men, you know that this works on numbers and statistics, what does that mean? Larger samples will give you more confidence. There's more samples. What does that mean? That means that the model you're gonna learn from this will reject qualified females at a higher rate than it will reject qualified males. That's simple, it's just statistics, okay? Now if you don't do anything about it, we'll get to that at the end. Final example of breath testing here. Let's say government wants to get peak, of course if you do a random testing to begin with, so you survey a variety of areas in the city, you randomly pick people. But the goal is to learn from that, to use a machine learning tool to discover what are the patterns of drunk drivers. Where they live and what is their characteristics, their demographic, you need to learn, because your objective is the next time you do it, you pick the people who are drunk, you don't want to be stopping people who were not drunk all the time. You need to do it randomly at the beginning, but at some point hopefully if you have a clever AI, you should learn who are the real drunk drivers, just go for them right? Okay, let's say you do that, it makes a lot of sense doesn't it? What could go wrong? I mean there are many traits that people, of their personality, peoples age, peoples gender, peoples race, they are not things you can be held accountable for, and yet will definitely be correlated with the outcomes of these tests. So what the machine will learn is just a survey of very specific kind of demographic, whatever that kind turns out to be. And because you don't stop thousands of cars, you stop hundreds of cars, in the top of that list who will turn out? It will be people from exactly the same demographic, exactly the same neighbourhoods, exactly the same profile. Alright? And this is all using AI off the shelf as we know it today. So what can go wrong? So let's review now to close the presentation. Your description of your goal needs to be correct. There are two things in this delegation framework that you need to give the machine. You need to give a definition of the goal, and you need to give the data that the machine uses in order to learn how to best achieve the goal. If you haven't specified the goal in the right way, that really reflects what matters about your problem, you could have an issue. If the data that you were using to train the algorithm reflects some historical facts that you don't want to perpetuate into the future, that can be a problem as well. And guess what, how often do you think we have these situations, these two problems? How often do you think you actually specify the problem precisely? How often do you think the data is perfectly reflecting how the future that humanity would like to have is there? Never, never okay. So basically you need to think about this from the ground up, it needs to be really ethics first and AI later. I've said this a few times, and I'll say it again, you should learn about AI, but not faster than you learn about ethics. Alright so this can happen in all fields as I just said, it's always the case that if you're building software using AI, you specify a problem, you specify it wrong, you get a data, the data reflects a path you don't want to propagate into the future, you've got issues. And there has been a variety of high-profile cases on recruiting, criminal justice, people being sent to jail or basically denied parole because they happen to be black instead of white, and you can actually prove that. So this may actually sound fairly dark, but think about it, it can actually be completely the other way around. The fact that these machines have these side effects, the fact that AI the way it's designed today is creating all these other problems, what does that mean? Well who has built the AI, other AIS? No, AI is a technology, it's just like a bridge on a car, we build it, we decide how we build this technology. Why are we having these issues? Is it because the AI acquired some agency and its out there like the Terminator, crazy and trying to kill everyone? No, it's because we don't yet fully understand how to design AI that doesn't create these issues, its lack of knowledge. And how do you acquire the knowledge required to fix problems? Just one answer, research. You need to study, you need to think hard, you need to enlist everyone who wants to think hard about these problems and get them to work on them. And if government wants to solve the big problems and the big challenges that it has, what it needs to do is first to realize that these off-the-shelf tools that you have out there today, if they are deployed for the citizens that the government is serving, you are going to have many blind spots, there will be many issues that you won't see. Not like the regional issue with the power supply here at the beginning of the talk, not like the issue of the gorilla, because you can see those. The problems are insidious things that are happening and no one sees, and in order to detect them you need science, you need some sort of new science. So you can't deploy, second point, you can't go out and deploy these algorithms without actually understanding the potential risks involved in every situation. Because here's the good news, this science is not well developed, because if it was well developed, we wouldn't be here talking, I wouldn't be giving this talk today, I wouldn't be talking about issues because we wouldn't have issues, that science will already have translated to the technology. The problem is that the science of ethically aware AI isn't yet developed, but we are already starting, and that's the good news. The good news is that we have already discovered a number of things about this new science, but these things that we have discovered haven't yet translated into the tech, because they are still in the research domain, so what do we have to do? Two things, we need to transfer them as fast as possible to the tools that we are using to make the consequential decisions about people. That knowledge that we already have. And the second thing we need to do, is to continue relentlessly increasing our amount of knowledge about this science, create new knowledge. Because there are two games, create more knowledge and apply that knowledge, create more knowledge and keep on applying that knowledge, and those two things we need to keep doing. And that's why I say that government should focus on helping to deploy these systems, but with the advice of experts. And likewise, fund research to continue understanding more of this space. Now just to close, as I was introduced before, my name is Tiberio Caetano, I am the chief scientist of the Gradient Institute, the Gradient Institute has its founding partner, Data61, among AIG and the University of Sydney, and we are a new independent not-for-profit research institute devoted to build and make AI systems that are ethically aware, and develop this science of good AI delegation so to speak. And we research, we apply and we facilitate the adoption of these tools, to increase the likelihood that you will do something sensible as opposed to something stupid. And you will soon, I haven't put there, but as of March 2019, start to help train both technical people, but also decision-makers in organizations to deploy AI systems that are more ethically aware. And with that, I should conclude. Oops, let's go back. And that was my message for today, thank you very much.

- Thank you for repeatedly referring to my battery dying Tiberio, I appreciate it, that was fantastic, thank you. We have such a stellar line-up of people today, and the final speaker, last but certainly not least, is Audrey Lobo-Pulo, so Audrey, you cannot have a conversation with Audrey without just being blown away, so she's got really deep experience in the federal government, and her area of passion and interest, which she has taken into her new consultancy, is around how you adapt innovative new technology solutions in governments to optimize public policy outcomes for all. It's such a pleasure that you could make it today Audrey, thank you for coming.

- Thank you, I'm very excited to be here. I want to thank Pia's team, and Data61 for inviting me, I'm so excited to be here. And it is the last talk, so I'll try to keep it brief. And try and get through a few stories. So what I really want to talk about today, Katie briefly mentioned my background, I've had over 10 years of experience with social policy at the Federal Treasury, and am also very passionate about data and analytics. So I think social policy is the perfect platform for us to look into how AI and policy can mix together. So I want to start with a story. Last week I was in a bit of a hurry, I ran over to the station wanting to buy my ticket into the city, and there was a man in front of me at the Oval vending machine, and he was well dressed in a suit, he looked educated, and he was increasingly frustrated, because he was pushing buttons on this touchscreen, wasn't getting the ticket that he thought he had requested, and I could see the veins in his neck pulsing, and he started swearing at this machine. And he said, damn I hate these machines. Why can't we have people, what's wrong with having people at the desk. This takes 10 times longer. At this point I was sure I had missed my train, so I offered to help, which added to the frustrations. He got his ticket and he slammed the machine, kicked it and walked away. And so I was a bit frazzled, because it was my turn to buy my ticket. And I was also torn in this place, because having been in government, I know how much thought and effort we put into how we design the way we interact with people, but I was also a citizen, and I could empathize with his situation and frustrations of not being able to get what he thought was a basic service. Now this machine has no AI, but the reason I used this example, is because I really want to touch into the issue about emotional intelligence and how we interact with people even before AI comes into the picture. So that's a bit about the back story, and I want people to start thinking from a government perspective about how are we designing systems, how are we designing machines, and how are we interacting with our citizens. And how do we use technology to really reinforce a sense of well-being, rather than making people feel quite alien to the technology.

So I started digging in a bit about the research around what makes us human and what makes us machines. And I came across Danielle Krettek, who from last week onwards is now gonna be one of my great gurus, and she is actually the founder of Google's Empathy Lab, and what I find really interesting, is that Danielle's work actually feeds into a lot of the Google AI products. She's got a background in technology, film, art, architecture and social impact, and has worked with the likes of Nike, Apple and Google. And the really interesting thing about Danielle's work, is that she actually believes that feeling is really the future of AI. And I, having spent as much time as I have in social policy design, I really believe that we need to integrate what is essentially core to being human into the way we think about AI. So I know we've had a lot of background material about how AI works, and my mission today is to try and make AI less intimidating, because I want people to realize that what we bring to the table as being humans, is also, if not more important, than what AI has the potential to offer. So I'm of the view very similar to Danielle, that as technology improves and we have an increased amount of IQ with machines, we then need to step up and try to match that IQ with our EQ. I want to talk to you about a few stories, and the next part of this obviously is what we think about when we think about relationships between government and citizens? We think about is it purely a service delivery, or is it a relationship, and how do we quantify the quality of these relationships? Going back to some research that was done by John and Julie Gottman, I'm not sure if people in this room have heard about this. But these two researchers were able to predict with a very high level of accuracy the longevity and the quality of relationships in marriages, to the point where they could almost predict based on observations of how people interact, whether or not a marriage would end in a divorce or a separation. And to me that was really interesting, because what they found was that what determined the quality of a relationship were the number of positive interactions as a ratio of the negative interactions. And that magical number was about five to one. So if you had five times as many positive interactions as negative interactions with people, you could say that was a good relationship. Some of the best relationships they found, the ratio was about 20 positive interactions to one negative interaction, so this has got takeaway messages on all fronts in your life. But again, with respect to relationship between citizens and government, can we create an environment where our citizens feel five times more positive, or have five times as many positive experiences with government as negative experiences? And how do we think about this when designing for AI? Back to Danielle's work, I'm going to talk about two studies, and the first one I found extremely interesting. So Danielle worked with Google and tried to create a couple of virtual assistants. She was very keen on testing out this whole concept about empathy. So the first assistant, virtual assistant that she created was purely information driven, it gave users as much information as they needed to make the best decisions. The second one did not provide as much information, and sometimes there could have been more efficiencies, but that was not optimized. But what that was, what that assistant was able to do, was that it was actually emotionally attuned to the clients and the people using that assistant. And the third one was a bland neutral kind of assistant. And the interesting thing about this study is, what she found is that people were much more keen on using not the virtual assistant that provided the most efficient solutions and the most amount of information, but people actually gravitated towards the ones that were more emotionally attuned. And what was even more interesting, was that two months after the fact, when she went back to people to ask them, which virtual assistant that they decided to keep, she found that there was three times as much loyalty to the virtual assistant that was emotionally attuned. So I guess what I'm saying here is that there is something intrinsic about the way our biological processes are, and about the way we perceive the world, and the way we empathize with people, and the way our emotions are wired that makes us feel connected to other people. And is this essence and quality something that we can integrate to complement the AI services that we provide? There was another study, which was done by Harvard, and this one really blew me away. So they had a bunch of people, and these people were asked to sit in front of a computer, and the computer would ask them a question that was actually quite intimate. So the question they asked was what are you guilty of? And some people got frustrated, some people thought it was absolutely rubbish, and they walked away. And so that was the first part of the study. But then the interesting part was, they had the computer address the person as follows: it said something like, I am a computer, there is no person behind this conversation, but I do want you to know that I am prone to crashing, I have bad days just like humans do, and unfortunately sometimes you might get frustrated with the crashing. And to the researchers surprise, what they found was that people at that point were willing to divulge so many personal insights about themselves and what made them feel guilty. Which was very interesting, because they are both computers, but in one case, even though you were cognitively aware That this computer did not have empathy, was not able to really connect with you emotionally, people were still comfortable enough To divulge a lot more information. So I think that's really interesting when it comes to AI. Because when it comes to social policy, we are essentially dealing with people. And it's about making people feel safe, and how do we make people feel safe in an AI environment? The other classic example I like to think about when I think about AI, and I know Tiberio mentioned a little bit about the context of ethics, but the perception around the data that we collect and how we collect this data. So we might be talking about different communities, like indigenous communities. How do we manage AI processes over peoples lifetimes, what does it mean to have an AI interaction with someone who's just lost a relative, how do they interact with government. People have different stages in their lives, how do we track those moods? And in a computer sense, I know we have what we call an airplane mode for a phone, which sort of sets us into a different space, we have the night-time mode where the phone rings a lot quieter, but the question is, it's not just modes on a machine, it's also moods of people and the journeys that we go through in our lives. So how do we design AI in a way that is a lot more empathetic, and actually relate to people, and relate to people's humanness? So this is where my passion for open models really comes in, and in the spirit of transparency and open government, I'm very, very excited by the potential of algorithms being open to the public, so that as a collective we can engage and co-create our algorithms to better have these public discussions about biases in data, what they mean, whether or not decision making is purely about efficiency, or whether it also connects with people's human journeys along the way. So finally, I know we started the day about, Pia talked about AI and democracy and how we can create a better world, but my hope for you, and I think it's a fitting that it is the last talk today, is that I really hope that AI can enhance what makes us intrinsically human, to even better heights, and really define us, define our humanity, and not create a gap between people and machines. Thank you.

- So now we'll move onto our panel talk, and joining Audrey is the other people you've met already, Tiberio, Richard and Nathan. And joining them is also the director of our policy lab here at the FSI, Tim D'Souza who just fell down the stairs.

- Thank you, so given the limitations in time, we are going to open up the floor to questions straightaway, rather than Zach and I throwing questions, I think we've covered a lot in the hour and a half or so since we've been here. But to be honest we have a very cluey engaged audience here, which is fantastic. So we might go straight to questions. Does anybody have any questions? If you could just state where you're from that would be great thank you.

- [John] I'm Dr John Selby from McQuarrie University. My question relates to the discussion we had successfully on machine learning today, and machine learning is very good at finding correlations, but it doesn't have causative models. Machine learning algorithm is very good at finding patterns that may or may not exist in reality, but they may exist in the data. So what if it's you putting into going beyond just correlations to including causative models into your AIs, so that we can avoid hopefully, a number of the problems that you've identified?

- Who would like to start us off on that one? Richard.

- That's a very good question, and actually I was reluctant to put causality in my slides, maybe I should have put it. That's a very good reason actually to answer this kind of question. So what needs to be understood with causality, is essentially it has become a field of its own, and recognize that as such quite recently, in particular with the most prestigious award in computer science awarded to one of its key people. And so people, and we are part of this team, are starting to think about the way that causality can be included in models in such a way that if you think about correlation when you learn the model, it's not necessary sometimes, you may just be happy about what your model is learning without thinking causation, but we are putting some thoughts into understanding what's happening here. There's a lot of causality actually intersect a lot of other constraints like privacy. You may actually use causality to infer private things about people essentially. And so it's a very tricky question, it belongs actually to the trickiest ones, but we are putting some thoughts into that, and that's a very exciting field actually.

- A brief comment on that. So even though I've been working on machine learning for 20 years now, I've been working on causality for about seven years now. Because there are many problems, most problems that we are interested in this life. They are problems of causality. You need to understand what you need to do in order to obtain what you want, and that's the problem of causality. Now as Richard said, in the machine learning community, and in many other communities, economy, social sciences, health sciences, epidemiologists, in a variety of different communities have been working on this from a statistical data driven perspective for a while. There has been a lot of progress, of course it's the basic thing about science, the scientific method at its very foundation is about finding causal relationships between different things, between actions and what you want out of them. So I think that causality connects very deeply with ethics, because ethics is not just about what you want, your intention, ethics is about your competence to actually do what will cause what you want. I mean, let's say in the days of the witch hunts, the people who were hunting those witches, they were actually concerned that the reason why they had famine and plague and all that was because there were these six women dancing counter clockwise, so they performed the utilitarian calculus and went there and killed those women. So they had good intentions, because they wanted to save 10,000 in killing those six, so the ethics was flawed, because of a causal problem, they didn't understand basic notions of causal relationships all the natural phenomena that actually creates all those issues. So yes, it's a massive question, I've personally been working on it for about seven years now, and it ties very tightly into both ethics and machine learning.

- [Zach] Any more comments from the panel on that?

- I might briefly, I might briefly add that it's not just a machine learning issue, and there are many, many other methods that have been used in the social sciences, where causation is up in the air really. So take that with a bit of caution.

- Thank you. Next question? Yeah, please.

- [Man] Hi, I'm from Property New South Wales. Just had a question regarding, there's a lot of talk about AI, and a lot of fluff out there as well, in terms of implementation mode, particularly on areas such as smart cities, you see a lot of providers like IBM, Microsoft, Ericsson, what do you think the role of government should be in terms of partnering with external bodies, how can we get more into the implementation side, because there's a lot of talk, but I'm not sure about the actions?

- I'll take a stab at this one. Where the private sector has great ideas and great technology, we should not hesitate to leverage that. There are a lot of opportunities for that kind of partnership, especially where you can rely on the private sectors investment in their technology, but that doesn't mean the government shouldn't be looking to bring those skills in-house, bring that investment in house and work on its own technology. Ultimately when we are working on AI and machine learning, we are part of a community, it's not a problem that will be solved exclusively by the private sector, it will not be a problem that's solved exclusively by the public sector, we all have different motivations and intentions, but there will absolutely be a role for partnerships, but we shouldn't as government generally rely on those partnerships.

- Just a quick comment, I think we need to always remember what are the incentives behind whoever you are working with. At the end of the day it comes down to that. So sometimes those incentives will align very well, and you will get what you need, and they will get what they want. Sometimes they won't align well, and you need to develop the skill To judge when that will be the case. So I will add that the solution is not only the private sector, the solution is the private sector or the government. Or take it in house, there is a growing non-profit sector growing out there. Just like today we see all this movement for environmentalism and healthy eating and ethical ways of pursuing many goals, and to accomplish great things from an ethical non-profit viewpoint. The same is starting to happen on the technology front and the nature of open source software for example, was early days what I'm saying there. And it's starting to happen now in this world of AI as well. So there's an ecosystem, and you need to consider all the options, but you need to think very carefully and ask yourself this question, what are the incentives of the people you are working with, that's the fundamental question that always needs to be asked, and often those incentives will align well, but sometimes they will not align well, and that's the time to look for an alternative solution.

- I think I might just jump in here, representing the public sector side of the equation. Look I think in my experience, the value that the public sector, all the people in this room can bring to these problems, is defining the outcomes that we want. And then leveraging the expertise that exists in the private sector to help pick apart those problems, make sure that we are not driving unintended outcomes, and to partner and leverage that expertise, and then engage in that knowledge transfer as well. Because just in the way that almost every organization is now a technology organization, we are all dependent on technology to deliver the outcomes that our organizations need us to, AI is just gonna become another tool in the kit, and we need to understand how those tools work at a decision-making level right throughout our organizations, from the top all the way down to the working level, and that's only gonna happen if we do leverage the expertise that the private sector offers us. So looking for those examples where the synergies and the desired outcome are really strong, and pursuing those. And there's lots of opportunities out there to do that at all scales. One of the things that surprised me was, again being able to tap in, from really quite a small scale exercise into IBM, that's huge power that we can leverage at minimal cost.

- And it's worth noting as you detailed to some great extent, that this is already happening, these partnerships are already ongoing now, and new partnerships are being built, so beyond what Revenue's been doing, there are things like the Western Sydney infrastructure strategy, which is being developed with private sector developers, with local councils, with various agencies, there are partnerships between state and federal governments, the Policy Lab is also partnering with Gradient Institute in the following weeks and months to work on questions of ethical AI, and how we build frameworks to enable the whole of government to on-board AI in a way that would be ethical by default. So yeah, it's part of an ecosystem, and there are interconnections all the way through it.

- Just one quick remark on this one, in every system, this starts with the skateboard to a car and a space station, you have a set of technical debts that are associated to the system, how it was built conceived, eventually flawed, you have additional technical debts for every system that uses machine learning, and it's important in my point to keep everything and everybody in the equation, because not one system is going to fit all constraints and all purpose, that's very important.

- Can I just ask on this topic, is there a community of practice within New South Wales government. I know that at the data sharing event, Kate Cumming had mentioned that there are COPs get together and share best practice around government. But I know the Department of Education, we are working on a machine learning transport for New South Wales, very active in the machine learning space. There are quite a few agencies that are grappling with similar problems, is there a community practice or some sort of mechanism to help? Does anybody know?

- There's a new policy community of practice for New South Wales government that started just before Christmas, so that may well be the avenue you're talking about.

- That could be a good forum.

- Okay any more questions, we had one in the middle there?

- [Woman] Thank you, I'm a bit of a newbie in the space but that's alright. When you were talking the previous question, you were talking about causation and correlations and things, I'm thinking from my background as an engineer and computer scientist, who actually ended up doing human factors. And I actually think what machine learning is doing simply correlations from all the data that we are getting, the example you gave of the company wanting to maximize profits, well that is actually their goal, so they are gonna keep doing that correlation. And we are talking about stereotypes. So how are we gonna get the ethics and the trust part back into all the parts of business, which is what you're aiming for from researchers perspective. How do we do that so that we do go forward and don't end up in a space where the industry and the drivers are driving that. Because I'm not sure that causation is the answer that we actually want, or people will want when they are using their AI, they are going to actually want those correlations, and then take them, and the stereotypes just keep continuing to be in place.

- Happy to start. It's a very good question. My view as an open government advocate, is that trust is a very difficult concept. I see openness, and open government is the first step to transparency. An algorithm could be open, but it may not be transparent, because people don't understand what it means, and it's very difficult to understand the implications of what that algorithm means. So I'm of the view that that is just the first step in this piece is opening algorithms. The transparency part, which is how do we communicate to the average person, and how do we communicate to citizens and the public, what the implications of that algorithm has. And it's a question I think we will as government, and I'm not government, but I've only just stopped being government, we are grappling with, and we will grapple with in the future, do we do a social impact study of algorithms, to we look at devices and disclose bias of the data as metadata or as information around what these algorithms could potentially result in. How do we deal with that? So it is a very open question, and it's a very important question, but my view is that open government is the very first step in addressing this issue. Because yes, there will be actors and private organizations that have their own incentives and their own goals, but from a government perspective, if we are co-designing policy with citizens, and we have the spirit of open government, then it's a discussion that needs to be had.

- Government has very different context and very different drivers to the private sector. So yes, your large-scale grocery retailer is out there trying to maximize profits. We have a different objective, our objective is to improve the lives of citizens, improve the lives of residents. So we get to engineer with that goal in mind. So we can do things like achieve algorithmic transparency, we can tell the story of what the decision we are making is, and the criteria that go into making that decision. More than that, we have the opportunity to use AI to design the optimistic future, rather than just automating what we currently have. So we can say, okay we are making administrative decisions we are automating the making of that administrative decision. Now, there are set criteria that you're supposed to take into account when you're making that decision. Humans bring their own biases with them, so we can sometimes end up in a place where people making an administrative decision have brought in random criteria that shouldn't factor in. We can use AI, we can use machine learning to enable decisions that are made in a way that never takes in an extraneous consideration, that only brings in the matters that should be considered. We can have that decision documented, we can have the automated process set out which criteria it brought into the decision and why, and we can have that entire process be transparent to the person that it affects. Then, if they're in a position where they think the decision is still wrong, they are better able to appeal it, and perhaps the appeal is valid, perhaps it's not, but either way, we have the entire process documented from end to end, so it's easier for everybody, and you still have the level of natural justice that is actually being provided at an improved position than it would be if it was just being facilitated by humans so we have the opportunity here to create better processes, and not merely automate existing ones. But it comes back to designing with that intent in mind, of improving services.

- Yeah, I think I'd absolutely agree with Tim there. I think what is a stereotype? It is a cognitive shortcut that we are programmed to make for snap judgements, it's an evolutionary trait that we have developed over millions of years to respond quickly to things in the environment. And it is based on correlation. It's not based on the underlying causation factor. So I think the power that AI can give us, particularly when it's coupled with big data and the availability of big data, is actually driving really personalized and tailored outcomes to individuals in the government. So that we are actually coming out with better outcomes more tailored to specific individuals, because we are only limiting the considerations to things that should be considered and are relevant to the outcome that we want to drive.

- But to more directly answer your question, how do we avoid these undesirable outcomes, we do it in the same way that we do with any technology. We design with the intent in mind, we prototype, we test, and then once we've established that it's going to work, and it's not going to have undesirable outcomes, we scale, we test again, we scale again, we test again. The answer is simple, it's just not easy.

- Okay, I think we've got time for perhaps one or two more questions, anybody over on this side please, at the back there? Thanks Jack.

- [Man] Hi, I really liked your speeches by the way, I thought they were great. One question I had was about the unintended consequences of designing AI, and that led to the idea of, sorry just give me a sec, Asimov's Law, whether or not that would be something, a set of criteria or laws would be factored into an AI to protect the user and the host AI itself at some points, like a robotics law, as well as improving, or maybe using deduction and induction through the scientific method to help stream its decision-making, and whether induction is even possible through AI at all. Thank you.

- I love that. I love philosophy. So you talk about induction and deduction and all that. So machine learning tries automating the process of induction, but any ideas of how to avoid the unintended consequences of AI are very welcome, because as I said before, this is not something that we have solved scientifically, so there is a reason why these unintended consequences are happening. It's not only that the science hasn't been transferred to the tools and the technology, but the science itself is in its early days. So we need to advance that, we need to put more resources, more money into the research to answer your question, because we don't really have excellent answers to your question yet. So we need to put more resources into fundamental research to understand what it means to create safer AI from the point of view of minimizing harm for society as a whole. I can't say too much more.

- Just two quick remarks, one is, how can I say that? I believe Asimov's laws were essentially carved in a system, I believe it's going to be, let's say the role of an external body to essentially figure out what needs to be, or what can be done with artificial intelligence, as I said in my slides, machine learning people need domain experts from the outside actually to work out the problems and the constraints that are now linked with machine learning. It will be a design problem that belongs not just to machine learning, but also people from the proper regulatory bodies and government as well, not to limit the way AI can be developed, but essentially control the way it's going to be used. And another quick thing is that when talking about an unexpected consequences, that is so true that in fact, it was recently shown that you can have a system which is just flawless for the problem you want to solve, but somehow you forget that your decisions have impact, and the impact of your decisions are essentially going to make this very same system in the future totally brittle, so let's say the analysers of the system to precisely take into account these unexpected consequences need to think far-fetched essentially, to cope with this kind of very important question.

- I'd like to bring it back to government just quickly. If government wants to exploit this opportunity, use this new technology, we have to reframe the way we're thinking about it. Simply put, we have to be willing to experiment, we have to create a space where we can try out new ideas, and have them fail safely. And we need to be willing to do that until we get a better grasp of how we can use this technology. So that can go counter to a lot of the ways that we have done things in the past, but we need to be able to create a safe space to try new things and be willing to feel, and be willing to try new versions of those things. That is how we will better put ourselves in this position, and a lot of that is going to come down to funding the research that's needed. It's a new area of science, and somebody's got to fund it, and who better than us?

- Going back to the point I mentioned previously, I think the social impact side is really important too, and putting some of our effort and research into that aspect, in terms of which demographics would be impacted, by how do we assess risk, if decisions don't go according to plan, are also measures that we can take, to not safeguard, but to try to mitigate part of that issue.

- Zach, are we out of time?

- Yes, so just to wrap up, I'd like to thank you all for coming here today, and thanks for Melissa and Jake and Katie for organizing today's event. And you'll be able to see these videos again on the digital.NSW website. But lastly I'd like you to all join me in thanking all of our speakers. Thank you very much.

- And thank you to everybody who is live streaming. I think next time we might provide an opportunity for questions to come through from that audience via Twitter or something, but the questions have been great, thank you very much for your participation today.