**Peter Chen:** 0:00

Learning a world model of other humans, other human drivers. Pedestrian cyclist behaviour gives them a much better ability to build a driving AI, so it absolutely makes sense. The same idea of understanding the environment through a lot of data so that you can better anticipate what actions you should take also makes sense in our robotics world, like, in particular, robotic manipulation In robotics, infants. Cost and speed are very important In robotics. You need your robots to be acting all the time. Like that means that you have to really optimise your AI models to give output very quickly and so the robots can take actions continuously.

**Craig Smith:** 0:42

AI might be the most important new computer technology ever. It's storming every industry and literally billions of dollars are being invested. So buckle up. The problem is that AI needs a lot of speed and processing power. So how do you compete without cost spiralling out of control? It's time to upgrade to the next generation of the cloud Oracle Cloud Infrastructure, or OCI. Oci is a single platform for your infrastructure, database application development and AI needs. Oci has four to eight times the bandwidth of other clouds, and offers one consistent price instead of variable regional pricing. And, of course, nobody does data better than Oracle. So now you can train your AI models at twice the speed and less than half the cost of other clouds. If you want to do more and spend less, like Uber 8x8 and Databricks Mosaic, take a free test drive of OCI at oracle com slash I on AI. That's E-Y-E-O-N-A-I all run together. Oracle.com slash I on AI. That's oracle.com slash I on AI. Hi, I'm Craig Smith and this is Eye on AI. In today's episode, we delve deep into the world of AI and robotics with Peter Chen, co-founder and CEO of Covariant, an industrial AI robotics company. Peter talks about building a universal foundation model that can operate different kinds of industrial robots across three continents. We also discuss the role of world models in predicting the outcomes of actions and their potential to generalise across various robotic applications. Whether it's navigating the complexities of warehouse automation or envisioning the future of robotics, Peter's insights are as transformative as the technology he's helping to create. I hope you find the conversation as engrossing as I did.

**Peter Chen:** 3:10

Craig, it's great to be here. Thank you so much for having me. My name is Peter Chen. I'm one of the co-founders, as well as CEO, of Covariant, a company that is focused on building foundation models for robotics. A bit of a personal history about myself. I was born and raised in China. I got into programming and computer science at a very young age. It always fascinates me how much intelligence you can get by programming software, which is basically instructing a computer to act intelligently, but there's always something left. What if the computers can learn from data? You have to instruct every single bit of the intelligence rule. That's what got me into PhD. After my undergrad at UC Berkeley, I started my PhD in AI with Professor Peter Beal at UC Berkeley, focusing on two areas that have both become extremely hot today. One area is reinforcement learning, the art of having machine learning models produce actions, and some of these actions lead to good consequences. Some of these actions lead to bad consequences. How can you have models learn from their own mistakes and successes? That's reinforcement learning. Another area of my PhD focus was generative models. My PhD advisor and I co-created and co-taught the first graduate level generative AI class at Berkeley back in 2017, 2018, a long time ago. Obviously, we have seen how those fields have really taken off in the last couple of years. Those ideas that were once very academic, cutting edge and unproven have become a much more commonplace concept that people interact with. If you go to chat GPT, it is both a generative model and a model that is aligned by reinforcement learning. It's amazing to see that arc of transformation of something that was cutting edge, unproven ideas, to something that now is everyday and is still continuing to accelerate the AI's development. Another bit of my personal background was I also spent time early on at OpenAI. I joined OpenAI fairly early. When I started working out at OpenAI, openai didn't even have an office. We were working out of Greg Proppman's apartment at the time. Obviously, openai has done amazingly well and has really powered a lot of the AI revolution that we are seeing today. Maybe the thing that I want to call out that's super interesting to me, especially as I reflect back, is that some of the core philosophies that OpenAI started out with are really the same set of philosophies that are powering the success of OpenAI that we are seeing today. If I were to summarise the early research philosophies at OpenAI in its first one or two years of existence as a company. It was this belief of a foundation model, scaling up big models on large, diverse datasets, this belief in generative models, using generative models to absorb a lot of unlabeled data, unstructured data. Third is reinforcement learning, like the ability to teach agents or models, the ability to take actions in the world. Those philosophies heavily influence me and heavily influence what we do at covariant Also. It's the same driving force of what powers OpenAI success today. Fast forward to how we founded covariant. A couple of the founders at covariant left OpenAI in late 2017 to start covariant. We started covariant really with much of the same thesis of what powers the current success of the large language model. We believe in a single model. We believe in a single large model. That is a foundation model, which means it's a model that is trained on multiple types of tasks. It can leverage the transfer learnings across multiple kinds of tasks and have emergent behaviour. It generalised to new tasks better, but also performed better at any specific task than a bespoke model that is only trained on that task. And we had an incredibly strong conviction that this foundation model for robotics has to be the way to go to solve robotics problems Because, like obviously, we have seen the success of this foundation model approach for language, but the reason that it makes even more sense for robotics is that there's only one physical world, unlike in the language. Like where you're trying to compress the whole world of human knowledge, which includes many things that have nothing in common with each other, like what is the soil composition on the moon versus how do you play chess Okay, like both of these are knowledge that you can find on the internet, but they have absolutely nothing in common with each other and you're trying to compress all these things into one model. But if you think about building a foundation model for robotics like the robot could have different bodies and the robots might be doing different things, like it might be interacting in different kinds of environments. However, like all of these robots live and operate in the same physical world, and so it makes a lot of sense to build one foundation model that can learn from all of these different robot experiences and really understand physics and understand how you control robots to move in the world around us. So that was a little bit of the founding story of Covarium, and fast forward to today. Covarium has commercialised the first robotic foundation model that's ever built Like, so essentially one single model that is powering robots working in production in customer environments in three different continents, dozens of different robot hardware bodies. That same single foundation model is powering and solving problems in a lot of different industries. We're starting out from warehouses, but our long-term goal is to build this foundation model that can solve robotic manipulation problems in general, like across multiple other industries. So that's a quick introduction about myself, where I come from, like the philosophies and the technical ideas that have heavily influenced us and where we have taken it so far.

**Craig Smith:** 9:53

Yeah, I have a couple of questions. One when you were doing the first course on generative AI at Berkeley, was that in response to the development of the transformer algorithm, or did you integrate the transformer algorithm into that course, Because that has really accelerated it is today the core of generative AI.

**Peter Chen:** 10:22

Yeah, I would say Transformers was not a key focus of that course. So when we think about generative AI, you can think about it as, like the two major components of it. So one major component is what is model architecture? So what is your brain structure? So, obviously, a better brain structure can allow you to learn better, and so a transformer is a really flexible brain structure that can allow you to absorb a lot of knowledge patterns from the data. There's another side of how you train a generative model. That is like, how do you teach it? So, even if you have a very flexible brain structure, like, how do you actually give that learning? Like, think of it as the curriculum teaching methods. So in this case, it would be the statistical models that you use, like different versions of it would be the most popular one now, obviously it's diffusion or to regressive model of next token prediction, and then, like a little bit earlier ago, like there will be GANs, vaes and these different models that are like the statistical representation, like the statistical representations that you impose on the world, like which you can see as like methods of how you teach the models. So I would say, like the initial class, focus more on things like what are the different models that you can use, as opposed to a more focus on things like what is the underlying brain structure? Like, but you can like, swap and plug in with these kinds of things Like. So, for example, the idea of diffusion like you can implement diffusion with both convolutional new nets. You could implement it with a transformer based architecture and like obviously the currently the more popular one, like stable diffusion is implemented with a combination of convolutional new nets as well as transformers. You can really take the best of both worlds.

**Craig Smith:** 12:21

Yeah, the current. I mean there's been a lot of talk, at least in the literature, in the last few months about using large models, large language models or pre-trained transformer models, as agents. And then, more recently you and I spoke about this the other day Jan Lacoon and Alex Kendall at a company called Wave AI are working with world models and you know that that learn causality directly from sensory inputs, not through the filter of language, and that world model idea speaks to me because it seems much closer to how the brain learns, initially In the covariant foundational model. Can you talk about the architecture and how it works? I know that it's certainly proprietary, but then talk about these new developments with large language models and world models and whether you're integrating those ideas or whether you see promise in them?

**Peter Chen:** 14:03

Yeah. So those are really good questions. Maybe we would tackle them separately, like the idea of role models and then also the idea of agents in the language world. So, first of all, like this idea of learning a role model for any kind of what we call embodied agents makes a lot of sense, right, like so, like if you, if the goal of an agent is to understand the physical world and take actions in it and your actions have consequences, then you should have understanding of what that consequences are right, as opposed to just blindly trying things and say, oh yeah, pushing button A is better than pushing button B. Well, that's it. I mean, that kind of works if you have a lot of data, but that's kind of like a very naive understanding of the world. Like, if I push a lot of button A, it tends to give me better outcome than pushing button B. Okay, then I do push button A more. That's kind of like not having a very sophisticated understanding of the world and it's also not very generalizable. Like what if I like, instead of presenting button A and B to you, I give you a keyboard and you need to type in, like passkey, and it does different things. You really cannot take any of the learnings that you get from pushing button A is better, like if now you suddenly have a different way of interacting in the world, and so the idea, like the general idea of building a role model, is can we build agents that really understand the environment and have the ability to anticipate what are the consequences of the actions that it takes? And this idea has many different incarnations, like with how, what Alex and the team is doing at Wave. A lot of it is about anticipating other agents' behaviour. Right, like if you drive a car, like slowly edging it to pedestrians, like most often people would try to step away from the car if they notice the car is approaching them. And if they don't step away, like that means maybe they didn't notice that the car is approaching them. So there's like some interesting interaction that you can learn by anticipating other agents. Behaviour in that physical world, which is absolutely the most core problem to solve in self driving, is like this kind of multi agent interaction and the behaviour that you need to generate from there. And so by learning a role model of other humans, other human drivers, pedestrian cyclist behaviour gives them a much better ability to build a driving AI. So it absolutely makes sense. And the same idea of understanding the environment through a lot of data so that you can better anticipate, like, what actions you should take also makes sense in our robotics world, in particular, robotic manipulation. So in our case, like the world that you're learning is not another human being's what's in another human being's mind, but it's more like what happens to physics, like if I pick things up in different ways, like what is more stable, what is less stable, like if I throw things away in a certain manner, like where would it land if I want to carefully type things together, and so like again, like you have two ways that you can build this AI. Like one way they can build this AI is I just do a lot of blind, random trials and I see what happens to work and I just keep doing that. Or you could actually learn a sophisticated understanding of okay, like if I pick things up this way, this is a pretty stable way of grasping a certain item. If I pick things up another way, oh wow, this is like a very unstable, very precarious way of picking up an item, but I can, by anticipating what would happen in the physical world, like it gives the AI a much better ability to act in it. So we are a strong believer in this idea of a role model like, essentially, this AI that can learn about the environment and also anticipate what would happen to it. And then another thing, that it's not just more like you said, it's not just a more sensible way for an intelligent entity to learn, because that's more like how humans would learn. You don't just randomly try, you anticipate, like how things are, what would be the consequence of the action that you take. But, in addition to that, like those and I'm really amazing property about role model that we believe is under talked about and which is this idea that if you formulate the right role model that unified all robotic applications, right, so like you could have like, for example, like if you think about like a robot that is folding laundry, as opposed to like versus a robot that is packing a customer order in a warehouse, like at a service level, like there's kind of nothing in common with these two robots. Like robots. Like one robot is trying to carefully think about, okay, like how do I pick up a deformable piece of a t- shirt and how do I flatten it and how do I fold it? Another robot in a warehouse would be thinking about it. I need to pick up the item and find where the buckle is, and I need to scan the buckle. Like from a policy or from an action perspective. There's really nothing in common between these two robots, but what is in common about them is there's only one physical world that's powering them right. Like so, if you're learning a role model that understands, if I interact with the world in a certain way, what would happen in the next few seconds physically. Like that concept is universal, right, and so like what this is like, what makes this role model idea so powerful? Like is it gives you this formulation, or quite like an interface that is the same across all robots, like, no matter what kind of tasks they are doing, no matter what kinds of hardware they are using. It's the same, it's the same role model because there's only one physical world. Like now, suddenly you find a way to really scale up the data that can go into training robotic foundation models. Like, like for all of these foundation models, like one of the like. Obviously you need the right model, you need the right algorithm, but you also need a lot of data of the right kind, right like. So the role model actually opens up the possibility of training a large foundation model that's learning on a lot of data, because now you can pull the data from many different robots together like and like. It doesn't matter what kind of tasks they are doing and the environment that they're interacting with, there's the same set of physical principles behind them and that's what the model can learn.

**Craig Smith:** 20:48

And the initial training. That's sort of the ongoing training. But the initial training you can simply train from video. Is that right that shows the laws of physics or you know causality and that sort of thing, and then, yeah, then there's ongoing training. Is the network computers learning?

**Peter Chen:** 21:19

Yeah, so we believe video is a super important format for this. Like, how can you like there's just so much data that's encoded in video, but what we have found is that pure video is also not sufficient. Like, for example, like, if you just go to YouTube and just watch a whole bunch of videos, you only get a very partial view of the world, like. So they are like that. Let me just point out like two things that are missing in in like, the first thing that is missing is, in a lot of the cases, you don't really know what are the actions that are taken, like because you're just passively observing things and so you don't really know, like, what are the actions that are taken. And then the second thing is, like, in a lot of these videos out there, you also don't have a very detailed. It also lacks the very small details that are really important to robotics like. So, for example, like what is the actual velocity of a certain value down to a very fine degree of precision, like? That information is kind of hard to infer from a video like, but if you are controlling a full robot system like, you actually can get it, maybe from the motor encoder, but you can get the information in a much more precise way. In a lot of those, in a lot of robotics cases, you do need a pretty precise understanding of the world that video doesn't fully communicate like. So both the lack of understanding of what are the actions that were taken in videos, as well as the position that is required, makes this kind of what we call videos in the wild a useful source of data, but it's definitely not a sufficient set of data.

**Craig Smith:** 23:10

And then, where do you get it? Is that why that data is supplemented with data from robots operating in different scenarios? Or are there other kinds of data? Can you use synthetic data, for example?

**Peter Chen:** 23:31

Yeah. So let me maybe answer the broader question like so, when we think about how do you build a robotic foundation model, like a truly universal AI that can be powering any robot hardware to do any arbitrary things to a very high level of autonomy? We believe the data recipe for those three pillars, like the first pillar, is what we talk about. Like essentially, data on the Internet, like video data, image data on the Internet. Second thing, that second thing about it is synthetic data, like generated data that are not made. They may not look exactly like the real world, but they contain useful structure about the world that can teach the AI and you can get lots of interesting combinations of known factors or variations through simulation, through this kind of synthetic data. Like. We believe that's very important. But these two things like this kind of Data in the wild on the internet and synthetic data are both very useful learning sources, but in our experience they are not sufficient. Like they still lack the kind of actual Detail interaction with the world, like understanding cause and effects, understanding them to a very high degree of fidelity, like those are not present in these two datasets that we talked about. Like to give an example of like where the synthetic data breakdown, it's very difficult to simulate contact and anything that deforms, and so those are the kind of places that like okay, I can use in the way simulated data to Simulate something that's rigid or maybe doesn't involve a lot of contact, but as soon as you involve that like your simulations, quality or the precision of it would quickly decrease. So, at the end of the day, like we, what we have found is that the third bucket of the data that you need is robots Interacting with objects in the real world at scale like so. These are like the three data buckets that Go into training, a robotic foundation, models like so they don't on the internet, on the, in the wild, synthetic data and then large volume of robotics data's interacting with the real world in production and, and this is really the core focus of covariant. Ever since we started the company, like, we strongly believe in the idea that, in order to build the best robot AI, you need to have the most amount of robotics data that are the highest quality like, which is why we really focus on one Solving customer problems, like making sure we build a technology that is not just Interesting lab demo but it's something that actually works reliably 24-7 in an industrial environment and the robots are so reliable, so autonomous and deliver as such a level of throughput that Our customers facilities just completely depend on them. And once you have that, like, you have Robots out there that are generating data at an incredible rate because, like, we're deploying these type robots into Industrial warehouse facilities that process amazing volume. Like so, once you actually make these systems generate commercial value, you can collect Tremendous amount of data while you Generate value for your customers. So that's being very customer value focused, being very production deployment focused, is like one thing that we have really focused on as a company. The second thing that we really focus on as a company is collecting the right kind of data. Like. So it's not just about, oh, get the robots out there and then, like, as they get used 24-7, a lot of data gets generated. We also spend a lot of time thinking about, like, what is exactly the right kind of data that you need to collect from the fleet of robots. I'm out there so that you can actually enable learning, and there is lots of deep research, thinking and iterations that go into it, and it's still something that we are obviously very actively iterating on.

**Craig Smith:** 27:36

Yeah, and the architecture of your, of your world, of your foundation model is? Is it? Is it like a JEPA architecture that Yanlacoon talks about, where the model is encoding the data into a higher representation space and then operating in that space to make predictions? Or or or is it More along the lines of, you know, a generative, pre-trained transformer model, where Everything's being tokenized and you're predicting the next in a series? Well, I guess that wouldn't be a world model, but when you combine these things, yeah, there are.

**Peter Chen:** 28:37

There are definitely different schools of thoughts on how you represent this type of thing. Like one, one way to do it is like you can say like there's some explicit Layton representation of the world that I learned and is in that some kind of latent description of the world that I learned to predict, I learned causality. There's another version of the world like, which is, if you think about A large language model like that, just a transformer. Looking at all the previous words and I predict the next word. Like there's never an explicit Representation of a latent structure somewhere right. Like there's no saying, oh, I encode all of my previous words into like some latent space and I decode the word From it, but instead let you just have this large structure that looks at everything that you have seen in an auto, aggressively predicting the next one. I wouldn't say I don't. I think the jury is still out in terms of which one will be a more likely successful structure, and we can probably draw some Biological inspiration, or maybe draw some inspiration into the next word. And we can probably draw some biological inspiration or maybe draw some inspiration from how we work, like and you say well, like it seems more efficient to operate in some kind of latent space, like because, like then, you're operating in this, your reasoning about the world in a more abstract way, as opposed to looking at every single pixels and then trying to think about, like, what should the next pixel be? I would say like. Both are avenues that we look at, but we don't believe there's like one clear winner at this moment.

**Craig Smith:** 30:21

But you, you have a foundation model in operation. So what's the, what's the, the architecture? Can you describe how that foundation model works, whether it's Transformer based, yeah, just.

**Peter Chen:** 30:43

So I think there are really two questions in there, like one is like what does well, like what does role model look like? And like what are the other parts of foundation models like in the specific architectures that I use there? So the specific role model that we have, I would say it's more similar to the latent type of representation, like there's a more compact, more abstract representation of the world that's operating on and and we believe like that likely would continue to be the case. Like just because, like operating in pure pixels of images and videos, it's a pretty wasteful representation to look at. Like when you think about, oh, if I don't grab something in a stable way and it drops, like you in your head, you don't try to predict where each pixel would go right, like you have this high level notion of oh, something would drop. So we think, likely , that it will be more successful. But we are pretty open minded about it. In terms of the second question of what are the specific model architectures that are used in Our foundation models is a pretty wide set of architectures. Is not a pure transformer based architecture, like so it's not all attention blocks throughout. Like it's a combination of, it's a combination of convolution attention and, in some places, more structured type of attention, like graphical neural nets, like for specific places that make sense. So I would say, like the key insight there is in robotics, inference, cost and speed are very important. Right, because, unlike in a maybe a different World like, where it's okay for my like next sentence prediction to be a little bit slow, in robotics, like you need your robots to be acting all the time, like that means like you have to really optimised your models to give output very quickly and so the robots can take actions continuously. And because of that, we spend a lot of work on not just blindly following the biggest, most expressive architecture, but really trying to use the domain understanding that we have about robotics and really optimising the architecture to be more latency sensitive, like to be more compute, budget sensitive.

**Craig Smith:** 33:20

And you have to adapt this model to whatever hardware it's running on, and we talked before about the hardware constraints in robotics. Does the world model your foundation model? Does it for each specific robot that it's controlling? The robot has a goal or a policy, and does the model have to take into account the hardware configuration? Certainly, yeah.

**Peter Chen:** 34:07

So think about in the large language model world it's very common to use system prompts to configure the character, the tone or the styles of a language agent. In our world you would think of it as the equivalent of prompting to basically instruct a robotic foundation model to know, well, what kind of hardware am I using and what are the things that I can do with my current hardware body. So think of it as one base model, but then on top of that you add configurations or prompting that actually instruct the model on. Okay, you're now using this kind of hardware body and this is what you should do with it.

**Craig Smith:** 35:01

You've been working with this foundation model as a world model since the founding of covariant. Has the research moved significantly from when you started? From when you started? Oh, certainly yeah. I'm reading a lot now about LLM-based agents and all of this world model. I've followed Coons research and it's really progressed a lot. Certainly.

**Peter Chen:** 35:40

Yeah, there are many things that are accelerating at a very fast rate. One core thing is computers. When we started six years ago, the amount of computers that you could have access to is very different from the amount of computers I can have access to today. As computer availability goes up, you can train bigger models that have more expressivity. And then the second thing is data scale. When we started as a company, we had no production customers. Now we have robots running autonomously 24-7 on three different continents, so the data that we can generate is just at a completely different scale. And then the last one is that the field has also moved a lot very quickly. So, if you think about a lot of the really scaled up transformer architectures, how do you do large scale training? How do you do image generation by diffusion? A lot of things have happened in the last couple years that also enable us to build more sophisticated foundation models and world models, because I think the key thing to recognize is that a lot of these ideas that we are talking about like whether it's foundation model or world models, they have many different levels of potential expressivity. So, for example, the most rudimentary form of world model might only be able to allow you to predict whether I have successfully grasped a certain item or not. That is also a world model, but it's just a world model that's restricted to understanding whether I have successfully grasped an item. A much more expressive form of world model could be well, if I have a cylindrical object in front of me and if I push it a little bit, it would roll around, and if it's on a slanted surface in my rollback, that is a much more sophisticated kind of world model compared to a world model that is only tailor specific to one small use case. So I would say, because of the three forces that have happened, like the more compute that is available, the more data that is being generated by our production fleet and the AI field's advances have together allow us to build progressively more powerful foundation models, and part of that would be the world model. And these progressively more powerful models can allow existing applications to perform better but also open up the possibilities for newer types of applications. So I would say, in the world of building foundation models for robotics, we are seeing a very similar trend to what we are seeing in the large language model world. So you look at the difference between GPT-2 to GPT-3 to GPT-4. There are remarkable differences, like as you scale up compute data techniques and you get greater capabilities out of it, even though you can argue they are all the same idea of transformer plus next token prediction. But as you do those things, you get qualitatively different results and it enables orders of magnitude more applications.

**Craig Smith:** 38:55

Yeah, covariance robots, even though they're different form factors, operating controlled environments, relatively controlled environments, and the degree of randomness or unpredictable ability is within a fairly narrow margin. How long do you think before a world model controlling a robot or a foundation model wouldn't only be a world model that can really operate in a real-world environment, unstructured, uncontrolled?

**Peter Chen:** 39:40

It's a really good question. I wouldn't call the current environment fully structured, like the robots, like where the robots are operating in. If you think about all the objects that we see and manipulate in our day-to-day life, they go through a warehouse at some point. If the covariant brain-powered, what we call our foundation model, the covariant brain-powered robots operating in all warehouses, it's building a pretty sophisticated understanding of how you manipulate things, even in the fully unstructured world. I think maybe the question is getting more at when we can have robots that move around and are fully in the wild, as opposed to in confined space in industrial processes that are high volume? I think that mostly is going to become a hardware question as opposed to an AI question. I actually believe hardware is going to be the long pole in the tent, as opposed to whether you can build the AI layer that can navigate in a less structured world freely. There are a lot of companies and efforts working on this, like humanoid robots, maybe humanoid with wheels, maybe humanoid with legs or maybe different kinds of form factors that actually allow you to navigate freely in space and also do useful things with it. I think there are a lot of interesting hardware problems to be solved there.

**Craig Smith:** 41:19

Yeah, on the warehouse, for example, on the hardware side, we were talking about `Wave AI. The car is a robot and it's well refined. How long do you think before world models can be applied to autonomous vehicles with such success and stability that we can use them on the roads today? Or is that really a regulatory issue?

**Peter Chen:** 42:03

I think this question is much better answered by the self-driving car experts.

**Craig Smith:** 42:11

Let me ask a warehouse question. I know we're coming up to the end of the hour. I remember when I first met you guys and we were talking about the advantage of automated warehouses. One thing that fascinated me is, as you said, they operate 24-7. They can operate in low light environments. They don't need the air quality that humans need. I mean, do you have a vision for and, as you said, almost every object that we touch has come through a warehouse do you have a vision for warehouses of the future that maybe will be these vast underground spaces where robots are working tirelessly in the dark and a human only has to go down periodically when there's some glitch or some hardware problem?

**Peter Chen:** 43:22

I think it's going to be more of a continuum. It's not going to be fully light-soul. What is the extreme of that? The extreme of that is, you can think about the whole moon being colonised by robots. There's no human being on it, but there's a space factory there that keeps pumping out goods and they get sent to the earth autonomously. That's one end of the extreme. It truly lights out no human touch point. I think that's pretty far away. Ultimately we will get there, but what we believe is more gradual adoption and you can already start adopting this type of technology today. For a long time this form of technology would be adopted in the form of human augmentation. I have to stand there and pump out 500 units of goods a day. I can now supervise 10 robots that are producing 5,000, 6,000, 7,000 units of goods. Today I have a much more engaging job of overseeing 10 robots and looking at where they get stuck. How can I arrange the upstream goods coming in, organise it in a way that allows me to get more output out of this fleet of robots, and figuring out how can I unblock the robots as quickly as possible as opposed to just doing the same motion again and again like 8 hours a day. This form of augmentation we see as a way to solve the labour challenges that our customers have. It allows them to do much more with a much smaller pool of labour, as well as have a much better human experience for people that are working in this vitally important societal infrastructure. Now you actually have a job that is a lot more fun, engaging and also way more productive than before. Over time, as the technology becomes better, you can see that ratio chips improving. Maybe now it's like one person overseeing a fleet of 10 robots. In the future it would be 50, 100. At some point you would have a whole factory of robots and maybe just one person working around in it. Maybe at some point in the future we would get to that oh, there's like 2 billion space factories on the moon and no one needs to touch that.

**Craig Smith:** 45:51

Yeah, okay. Well, let me ask one more question, because when we first were talking during the pandemic, there was a lot I don't know whether it was you or one of your partners I had asked whether robots could operate a chicken processing factory. I mean, certainly there is some automation there, but there was so much death from COVID on these production lines, but the hardware wasn't versatile enough at that point to be able to deal with something as soft and floppy as a chicken carcass. Is that? Are we on the road to that? And then my final, final question is do you feel you talked about a continuum? Are we close to a step change in robotics or are we still on just a pretty steady incline?

**Peter Chen:** 47:10

So I don't actually have context on the chicken processing question, so I would maybe not answer that specifically. But on the idea of making progress on manipulating more dexterous objects, like deformable objects, in a more dexterous way, the answer is definitely yes. So when we look at our systems and the capabilities that they have, it's definitely understanding the physical world in a more and more nuanced way and in a more and more expressive way, and so, like we're on a way there. And then, in terms of the questions of whether we see a step change in robotics, I will say the most honest answer is we don't fully know. Like there are some forces that are working for it and then there are some forces that are working against it. So let me tell you the forces that are working for the huge inflection acceleration, the forces that are working for it is we are getting to the point that you can really scale up compute and data to train really large robotic foundation models, and we have seen this kind of step change, like this phase transition of capabilities in the language world as you scale to a certain size of model, compute and data, and we expect the same thing to happen, right, like we expect, like the foundation models that power these robots to get significantly smarter, to get significantly more general. Like I mean, I cannot comment on what would happen outside of covariant, but at least at covariant, like we are seeing that coming. Like how do you scale up data compute model size significantly to get a much smarter model? The forces that are working against it is the adoption still needs to go through hardware and it still needs to go through an enterprise adoption process Like this is not like chat GBT, that you can like largely turn on the faucet and or maybe not turn on the faucet but like go to a website and then you can use it Like so like the adoption process would be slower, but I would say in terms of the capability leapfrog, like we think we are very close to a very clear phase transition.

**Craig Smith:** 49:28

AI might be the most important new computer technology ever. It's storming every industry and literally billions of dollars are being invested, so buckle up. The problem is that AI needs a lot of speed and processing power. So how do you compete without cost spiralling out of control? It's time to upgrade to the next generation of the cloud Oracle Cloud Infrastructure, or OCI. Oci is a single platform for your infrastructure, database application development and AI needs. Oci has four to eight times the bandwidth of other clouds, and offers one consistent price instead of variable regional pricing. And, of course, nobody does data better than Oracle. So now you can train your AI models at twice the speed and less than half the cost of other clouds. If you want to do more and spend less, like Uber eight by eight and Databricks Mosaic, take a free test drive of OCI at oracle com slash I on AI. That's E-Y-E-O-N-A-I all run together with Oracle.com slash I on AI. That's oracle.com slash I on AI. That's it for this episode. I want to thank Peter for his time. If you want to read a transcript of the conversation today, you can find one on our website E-Y-E-O-N-A-I. And remember the singularity may not be near, but AI is already changing your world, so pay attention.