**Richard Sutton:** 0:00

There isn't a science around that isn't profoundly influenced by the availability of massive computing power and just greater regular computing power. It's the story of our age. It's not just the story of AI. The idea is to leverage computation to make useful things and understand the mind. These, all these things need a lot of computation. It's the fact that computation is becoming more plentiful and cheaper exponentially, for on the order of 100 years and can be expected to continue going that way. It's like doubling every two years, now every 18 months, and that keeps happening 18 months after 18 months, after 18 months, and it means you double and you double and things get qualitatively different every decade. And that's happened for a long time, for many decades, and will happen more so in the future. So we have that to look forward to. I think it's what we really should mean when we say the singularity. The singularity is that we have this exploding it's a slow explosion of computer power, and that that is fundamentally changing things.

**Craig Smith:** 0:59

Hi, I'm Craig Smith and this is I on AI. In this episode, I speak with Richard Sutton, the father of reinforcement learning and professor at the University of Alberta. We discuss his cooperation with John Carmack on Keen, a startup that vows to reach artificial general intelligence by 2030. Richard also talked about the Alberta plan, his ambitious five-year research agenda focused on building embodied agents with the capability to learn and plan through interactions with their environment. Sutton provides insights into the current state of progress new algorithmic developments and tradeoffs between simulated and physical environments in training and the ultimate goal of creating AGI. I hope you find the conversation as amazing as I did, so why don't you start by introducing yourself? I assume people know who you are. I've had you on the podcast before, but, for those new listeners, tell us who you are, where you are, and then we'll talk about the Alberta plan, which I find pretty exciting.

**Richard Sutton:** 2:15

Thank you, craig. Well, I'm Richard Sutton, I'm a scientist, I've been studying artificial intelligence for 45 years a long time and I'm up in North at the University of Alberta in Canada and I'm a professor in the Computing Science Department and also I'm a researcher at Keen Technologies and I got lots of titles and sub-rolls. But basically I'm just trying to figure out how the mind works and I've tried to do it in a very broad and interdisciplinary way, reading all the different thinkers on the subject and addressed from the point of view of psychology and how the brain might work, as well as computing science.

**Craig Smith:** 3:00

I've read a number of the recent papers and I can see this thread developing and I don't know whether it's just that you're writing more and so the thoughts are more developed in print or whether they're developing in your mind. But from 2019, when you wrote the Bitter Lesson, you talked about the idea that it's really increasing computation and the driving a lot of things, a lot of progress. That kind of coincided with OpenAI's scaling of the transformer model. I talked to Ilya Sutskover and I asked him whether your essay had triggered their interest in scaling and he said no, it was coincidental. But first can we talk about that, about how scaling and the availability of computational resources in Moore's Law has driven a lot of what's happened in artificial intelligence research, almost more than novel algorithms.

**Richard Sutton:** 4:19

Well, I think the first thing to be aware of is it's been driving things that are not just artificial intelligence. It's been driving all the sciences and all the engineering developments in the world. There isn't a science around that isn't profoundly influenced by the availability of massive computing power and just greater regular computing power. It's the story of our age. It's not just the story of AI, it's not particularly the story of AI. Ai has always known that it needs computation. The idea is to leverage computation to make useful things and understand the mind. Yeah, now, it's true that those of us who are very interested in connectionless systems or distributed networks, nowadays just called neural networks not particularly good terms, so I always shudder a little bit when I use it but those of us that have been doing that have been doing learning. I think that learning is important for intelligence. These, all these things need a lot of computation, and so they are limited by the computation available at the time. Okay, so let's be. What is this thing? What is the Moore's Law? What's called Moore's Law? It's the fact that computation is becoming more plentiful and cheaper exponentially for on the order of 100 years and can be expected to continue going that way, so exponentially. It looks like doubling every two years, now every 18 months, and that keeps happening 18 months after 18 months, after 18 months, and it means you double and you double and things get qualitatively different every decade. And that's happened for a long time, for many decades, and will happen more so in the future. So we have that to look forward to. That will continue having a tremendous influence on everything that's done. On the other hand, it's just normal, it's just what you would expect and those of who worked on AI for a long time have just you know, expect and plan for. And now it's coming. But it's an exponential, so exponentials are self-similar, so that means they look the same at every point in time. Every year it's you're doubling in a year and a half and so it's an explosion. As every exponential is an explosion, it's sort of. I think it's what we really should mean when we say the singularity. The singularity is that we have this exploding. It's a slow explosion of computer power.

**Craig Smith:** 7:08

And I had a really interesting conversation almost a year ago with Aidan Gomez, who was on the team that designed the transformer algorithm at Google and he now has a startup co-coher. He's Canadian and he said an interesting thing that he believes it could have been almost any algorithm. It didn't have to be the transformer, that the community got behind the transformer, poured resources into it, continued to scale it and it was scalable. I mean, that was important, that it's a scalable architecture, but it didn't have to be the transformer. And that made me think of you, because so transformers, the way he described it, at its core it's a stack of multi-layer perceptrons. With attention, you scale it, feed it data and it learns to understand language, or at least seems to understand language, but it's got all these obvious limitations. I've been talking a lot over the last couple of years to Yamakuten about world models and that to me sounded like a much more exciting direction for general intelligence, because not all intelligence is contained in language, or at least, or even less so, in human text. And then I see you guys come along with the Alberta plan and that sounded even more exciting to me. So the Alberta plan you're building the ideas to build an agent, ultimately an embodied agent, that has a world model or can create a world model through interactions with its environment. How is that different from LeCun's approach? At a very basic level.

**Richard Sutton:** 9:33

Very basic level. The good is that they're a very similar idea. You look at the parts of his architecture and the parts of the architecture put forth in the Alberta plan. They line up one for one. We're trying to do the same thing, we're going about it slightly different and we could talk about that. But I think to just focus on the differences might even be to distract from the big message. The big message is that you have to have a goal and you have to have a model of the world, and then everything is driven by using that model to take action and to plan action at various levels of abstraction in order to achieve the goal. Okay, so to me, this is what intelligence is Understand the world, use your understanding to achieve your goals. I'd like to formulate the goals as a reward and I'm super comfortable with that. Other people sort of grudgingly accept rewards, even though it seems kind of low level, but it's a natural approach, I think. I think it almost makes more sense to people who aren't steep and deep learning.

**Craig Smith:** 10:50

One thing I found interesting in the roadmap that you laid out for the Alberta plan, you start with supervised learning. And why is that? Is it just because it's easy?

**Richard Sutton:** 11:04

Yeah, I guess we do in a sense because we want to focus on continual learning, which is sort of an obvious thing. What learning means is it has something that goes on at all times. But the first steps, getting continual learning with nonlinear networks, is still challenging, even for supervised learning, and so it's natural to start at the simplest possible case which involves the fewest other factors, and that's a supervised learning case. Yeah, it's funny. Let me just say a few words about that, because there's sort of been a fight through a struggle throughout the decades between supervised learning and reinforcement learning. There's only so much oxygen for learning methods and all the attention that's paid to supervised learning somewhat detracts from reinforcement learning. There's a bit of a friendly competition and supervised learning has always won the competition because supervised learning is so much more easy to put into practice and for people to use and it's sort of less ambitious. But it's really important and really those of us who do reinforcement learning and try to make whole agent architectures we are consumers of supervised learning outcomes. We will use them as components of our overall architecture. So we need them and we can work on them and we need to structure them for our purposes.

**Craig Smith:** 12:38

I saw one of your talks. You make a distinction between tools and AI agents, and supervised learning falls into the tool category. Can you sort of start and talk about the evolution of the Alberta plan and then present to listeners what it is in its simplest form, and that'll give me a structure on which to hang questions?

**Richard Sutton:** 13:07

The Alberta plan is an attempt to understand intelligence as primarily a learning phenomenon. It's something that comes to understand its environment and then drives the environment to achieve goals. So the first step in the Alberta plan is the structure between the agent, the environment and their interaction form. The interaction is not exchanging states, exchanging observations like sensors, visual touch, auditory it's all abstract to those particulars, but it's got to be genuine observations and not state because state we don't really have access to directly. So the principles, number one principle I'm trying to remember them as I speak, but number one principle is this agent-environment interaction is sacrosanct. And number two is that learning or everything is, we could say, continual. I think we say temporally uniform, temporally symmetric in the Alberta plan, which means that there are no special phases where you like training and test. There's just life goes on and on. You get rewards or you don't get the reward you want and you get your observations and there is no teacher other than rewards, pains and pleasures. And maybe I'm not getting the four principles right. But another important point is that you are going to be forming a model and so you're going to plan. It's both trial and error, learning directly from experience and learning a model and then planning with the model. Both of these are important part of intelligence. Okay, so that's the background. Then we outline the 12 steps and the 12 steps really start with. Let's have learning that is temporally uniform. Let's have metal learning, metal learning maybe I should stop on that for a moment. Metal learning means learning to learn, not just learning one function, but once you are continually learning, you're learning this and you're learning that, you get many, many experiences learning and you can get better at learning. You can use those repeated experience with repeatedly learning to make future learning episodes more efficient. So, as part of that, you learn representations, you learn features, you learn step sizes Okay, so continual learning and then all the algorithms, and once we add metal learning and continual learning, we have to, in supervised learning, that we extend that to reinforcement learning, which involves its own set of issues, more interesting temporal relationships, and I think, like the first six steps are crafting the basic algorithms of reinforcement learning, working through them again to be continual and meta, and then we start to bring in the challenging issues like learning off policy and learning models of the world, and then planning and the fun, just to jump to the end. The last step is about AI, ai, ai's, ai, intelligence augmentation where we combine computers, ai's, with our own minds to make our own minds stronger. Okay now, one of the key steps in there was off policy learning and learning a model of the world. Off policy learning means you want to be able to learn about things that you're not doing, or you're not because you're not doing all the way to completion. So even like to recognize an object, you look at the object and you say how would you? You have to define that in some objective way, and the best way to just do that is as a sub problem. So, yeah, maybe I'll just sort of stop there. The most interesting strategy, distinctive strategy by the Alberta plan is the pose, is that the mind works by posing sub problems for itself and then working on them. And it's not. It's sure it's got a main problem, which is to get reward, but it also has many thousands of sub problems. It's also working on simultaneously and since it's not behaving it cannot behave for all thousand problems at once it has to pick one problem, like perhaps the main problem, and behave according to that. So all the other things have to be able to learn from data. That's not exactly on what they would do, and this is called off policy learning and it's a key to learning to achieve auxiliary sub problems, and also it's a key to something called the horde architecture is?

**Craig Smith:** 18:27

is that where that comes in, when you break a problem down into multiple sub tasks, that you learn?

**Richard Sutton:** 18:38

I was one paper where we worked on that idea. We developed that idea. The horde is the horde of sub problems. Each demon in the horde which is it could be almost viewed like a single neuron in a neural network, as achieving, working towards a different task, trying to predict a different thing, or maybe trying to attain a different thing. It's the view of the mind as decentralized. There is one goal and everything is ultimately driven towards one goal. But still it's a useful structure to have different parts driving at. How did you get together?

**Craig Smith:** 19:17

with John Cormack. Was that primarily because you need the funding and it gives you a vehicle to raise capital? Oh seriously, I mean you know Jan Lacun's got meta behind him.

**Richard Sutton:** 19:31

Well, it's just not really comparable. John's company is great, but it's still like a $20 million company, which is plenty of money for what we want to do now. John and I got together because we had similar ideas about what was needed, and also what was not needed, to get to AI or AGI. Yeah, so I read an art newspaper article, an interview that John did down in Texas, and I just could see that he was thinking about the way I was, even though our backgrounds were quite different. So you thought of intelligence. You had to. There's a few principles that needed to be worked out rather than so. This isn't a huge program to write. It's a few principles. We have to figure those out Not that many, maybe 10,000 lines instead of 10 million lines of code. So it's easy to get it's relatively. It's still hard to get basic research funding in the world. It's easy to get funding towards applications of AI, large names models particularly. Anyway, I'm really enjoying working at Keen and being able to focus on the ideas, and it's a calm company. There's a lot of thinking involved, a lot of contemplation. There's also experiments and we're trying to get the engineering side of it is really important, but for me it's been really great just to be able to regroup my thoughts, but Keen is implementing the Alberta plan.

**Craig Smith:** 21:23

Is that right? I mean, that's the project.

**Richard Sutton:** 21:27

Well, the Alberta plan is a research plan. It's like a five-year research plan, and so research is something you don't implement. Research is something you conduct and it doesn't always end up the way you want.

**Craig Smith:** 21:41

Yeah, I wouldn't say implement, but the work you're doing at Keen is informed by the Alberta.

**Richard Sutton:** 21:52

Yeah, absolutely, I'm working on that.

**Craig Smith:** 21:53

And the end goal at Keen is to create the embodied intelligence described by the Alberta plan. You know it sounds very confident.

**Richard Sutton:** 22:05

Well, a plan is just a plan and you know I think it's a good chance that it will work out as planned. But you know a five year plan. You make another one after four or three years. Yeah, so I wouldn't presume to know how it's going to work out. But at the same time, we have to make you know, we have to make our bets. We have to think hard about it. Just knowing you know, we may well be right.

**Craig Smith:** 22:37

But you know, your work is primarily in reinforcement learning. You wrote the book on reinforcement learning, temporal difference learning and Lambda, and all of that Is this. I mean, this is. This seems a much more ambitious project. Is this? Was it the success of the transformer scaling that said well, you know, let's do that with RL. That's why. Why are these guys? You know, everyone's celebrating what they're doing, but there's much more to be done.

**Richard Sutton:** 23:20

No, no, what you're seeing in the Alberta plan is perhaps bigger than the book, but this has always been the plan. We've always, in AI, tried to understand all of the mind and reproduce it in computers, and so that's a that is a big, enormous ambition. That's what it's always been. So the large language models are a bit, a bit disappointing in some sense. I mean, it's really good that people are getting excited and people are wanting to learn about it, but but it's not it's. I don't envision that it's the direction that will be most productive to pursue. Now, you know who knows? What I do know is it's not the most direction that's useful for me to pursue. I must learn some actions and goals and how an agent can tell what's true and what's not true. All of those things are missing from large language models. So, no, I'm not, they're not really. What are they? What they are doing? That's important Is they're showing what you can do with computation and networks and learning and that you can get enormously complex things and you can incorporate a lot of data.

**Craig Smith:** 24:45

Just shows the power and be an interface between humans and whatever you end up creating, the agents you end up creating, you still need a language interface to communicate.

**Richard Sutton:** 25:04

Yeah, but I don't, I doubt that what we're doing with large language models. Oh, that's right, yeah.

**Craig Smith:** 25:11

And, in other words, the models that you want to build. The agents you want to build would learn language as part of the learning process.

**Richard Sutton:** 25:21

Yeah. So it's like we say language, language last, you know, not language first. With large language, models are language first. We just say large language last. Just as Jan McCoon says we need to do, you know, rat level intelligence and then cat level intelligence, and we have to get those figured out before we try to make them develop.

**Craig Smith:** 25:41

Where are you on the plan? I mean, you figured out reinforcement learning. You can build agents. There are various architectures for creating representations from various kinds of sensory input and at that representation level then you can plan efficiently. So where in all of that, are you in your research?

**Richard Sutton:** 26:16

Well, it's a little hard to explain non-technically, but you can say some things Certainly. You can say that the various steps are not done entirely sequentially. You're always looking for areas of opportunity where you can make an increment of progress and those could be, you know, on step 10, or they could be on step three. But you could also I could also try to be very rough and say that we're at about step four now. We are still doing things where we're changing the basic, underlying fundamental reinforcement learning algorithms. We are not done with that. We need more efficient algorithms and I'm excited about some of the changes. New ideas were developing recently about how that can be done.

**Craig Smith:** 27:10

Can you talk about those new ideas at all? And don't be afraid of being technical. I'm pretty adaptable. I can pretend like I understand and my audience likely will understand.

**Richard Sutton:** 27:23

Okay. Well, one of the big things is efficient off policy learning and the use of important sampling. Important sampling is where you see how likely you're to do things under your target and your behavior policies and you adjust the returns based on those, the ratios of those two. And for a long time I thought that was the only way to adjust the returns. But now the forward correction of the returns I think can be done by changing your expectations, so like if you're expecting a good thing to happen, expecting a good action to be taken, and then a different action was taken, a more exploratory action, so this is a deviation from your target policy, which would be more greedy. And one way to take into account the deviation from the target policy is to just say oh okay, now I've done something, not best, so I'm just going to adjust my level. Now you're going to expect a little, a little less, and there's a way, there's a systematic way of doing that. That's gives us a new way to handle the off falseness of of our returns, and so this gives a whole new family of algorithms. So that's exciting. Now for exciting, maybe mostly for me, I think maybe the most accessible direction of, of, of excitement, of novelty is in continual. So there's, I'm going to say a bunch of things, and to me they're all going to have the same solution Continual learning, meta learning, representation learning, learning to learn, learning how to generalize state, how to construct a state representation feature. Finding that whole thing is is is coming and it will be a kind of. It's just a new kind of way, a new kind of way of doing the learning in deep networks, and I call it dynamic learning nets. See, a dynamic learning nets have learning at three levels, whereas usually our neural networks only learn at one level. They learn the level of the weights and in addition, we also want to learn at the level of step sizes. So, all of every place you have a weight in your network, you also going to have a step size. So a step size is sometimes called a learning rate is much better to call the step size, because a learning rate will be influenced by many other things. So if we imagine a whole network, all these weights, next to each weight is a step size that is adjusted by an adaptive process that's adapted in a meta learning way, a metagradient way towards making the system learn better rather than just perform better at an instantaneous moment in time. Learning rates or step sizes don't affect the function. They don't affect some function implemented at a particular point in time. They don't affect with the network, does they affect what the network learns? And so if you can tune the step sizes, you also get learning to learn and learning to generalize. Well, and things like that. The last three, the last element that we wanted to have be adaptive weights, step sizes. The third one is the connection pattern, so who's connected to who? And so this will be done by an accretive process of like. Let's say, you start with a linear unit and it learns, say, a value, function or a policy. It does the best you can with the features available, and then it needs to induce the creation of new features, because you need to learn a nonlinear function of your original signals, and so you need to create new features that have become available to that linear unit and in this way you grow, in a sort of organic way, a system that can learn nonlinear functions. So this is just a different way of ending up with a deep network that was all learned, including all the features, dynamic learning, that's.

**Craig Smith:** 31:34

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 spiraling 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, 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 oraclecom slash ion AI. That's E-Y-E-O-N-A-I all run together. Oraclecom slash ion AI. Where is the data, the input data, coming from?

**Richard Sutton:** 32:55

Well, the input, data and reinforcement just comes from life, from doing things, seeing things Right. There is no labeled data set. Yeah, maybe I should have said this from the very beginning. The whole idea of I call it experiential AI is that no one makes you data. You grow up as a baby and you play with things and you see things and you do things, and that's the data. And the trick of reinforcement learning is how do you turn that kind of data into something you can learn from and grow a mind from? So the beauty and the limitation of supervised learning is, they say well, let's not worry about that for now. Let's assume that somehow we have a data set with labeled things and let's work on this sub problem. That's a great idea Work on a sub problem, figure it out and then move on to the next thing. But really we have to move on to the next thing. We have to worry about how the data set quote data set is automatically created from the training information. There isn't ever a data set. Data set is such a misleading term. It suggests that it's easy to have this thing and store this thing and curate this thing. Really, life is full of you do things, things happen, and then there's one everything is fleeting, you don't have a record of it, and it would be enormously complex and not only valuable to have a record of it. The feeling is totally different in reinforcement learning and supervised learning, and in, particularly, the way I would adjust it. Many people do reinforcement learning by creating a buffer or a record of all the experiences that have been retained, that have been occurred at least for some period of time, and I think that's an appealing, but it's not where the answer is. The answer is embracing the fleeting nature of data and making most of it when it happens and then letting it go.

**Craig Smith:** 35:02

That's why you want to make an embodied system, so that you have all the five senses or more as input data and in the experimental stage, are you using images or video?

**Richard Sutton:** 35:29

Well, you need, as you say, an embodied system, an interactive system that influences its input stream, its sensory stream, and then you get that interaction, and for a long period of time. You can do this in simulation or you can do it in robotics. I still know what's the best way, or if the best ways, do both, and maybe first one and then the other. John is interested in learning from video and his view of the experience is you have massive numbers of video streams. It's like you're viewing 500 channels of television and then you can switch to look at one, look at another one. Other people in Keen, my close colleague, joseph Modial. He's interested in robotics and he thinks the best way to get an appropriate data stream is to actually build robotic hardware. You know it's important that the world be large and complex, because the worlds we want to address are large and complex, and so you want things like video and you want large data streams. Now you can use simulations to generate even video streams, simulated video, but inevitably those simulated worlds are really quite simple. They have an underlying simplicity. They have objects, perhaps in three dimensional structure, maybe they're rigid objects, and vision is a very particular geometric form. They are generated and they are made up worlds and they are generated, so they're really. The worlds are less complex than the agent. Their goal would be to spend most of the computer power working on the mind and just a little bit to create the simulated data. And that's reversed the way. It really is right. Every person is, maybe has a complex brain, but their world is much more complex, not just because the world consists of all these physics and matter, but it also consists of other minds, other brains and other minds out there, and what goes on in their minds matters, and so the world is inherently vastly more complex than the agent, and we've reversed that when we work on simulated worlds, which is always concerning. Anyway, those are some of the issues in the trade-offs between working with simulations or with physical worlds.

**Craig Smith:** 38:09

Nonetheless, you need to develop the architecture and the algorithms before you worry about the data stream. I would think.

**Richard Sutton:** 38:19

Yeah, but you want to develop the right algorithms and if you're working with the world, it's not representative of your target world. It can be misleading, but you're right and that's what we strive to do. I don't know if you know, but I think of my own work. As almost always, I want to focus on some issues. So I make a really simple instance of that issue, like a five-state world, and I study the hell out of it, but I don't like try to take advantage of its smallness. I study algorithms that are in some sense even simpler than the simple world and I stress those algorithms and see what their abilities are. So it's always part of research as we simplify the world, understand it fully, just like a physicist might make a simplified world with a ball rolling down a ramp, and it's a really simple world and you try to eliminate the friction and you eliminate other weird effects.

**Craig Smith:** 39:18

Yeah, have you paid much attention to Alex Kendall's work at Wave AI? Do you know that company? It's an autonomous driving company. They have a world model called Gaia 1. And it's it's similar to what Jan Lacoon is doing. It encodes representations from video, from live video, and then plans based on those representations, and it can control a car from the representation space.

**Richard Sutton:** 40:00

It's actually pretty remarkable, so let's talk about the world model, and what kind of world model would be appropriate for autonomous driving. So let me say some things that are mistakes. They're a natural seeming but mistakes. In my opinion, the mistake would be to make like a physics model of the world, or to try to make something that could simulate the world and produce the video frames. You don't want the video frames of the future. That's not the way you think. Instead, you think oh, I could go to the market and maybe there would be strawberries. You're not creating a visual video. You're jumping to the market and then your strawberries could be different sizes and positions, and still there's not a video. There's an idea that will happen if you go to the market. So people have realized this, like Jan Lacoon used to talk about generating video of the future and then he realized it would be blurry, and now he realizes that you need to produce outcomes of your model that are not at all like video streams and not like observations at all. They're like constructed states that are the outcome of the action. Ok, so this is a very different from a partial, differential equation model of the world, and so it's very different from what self-driving car companies start with. Self-driving car companies start with physics and geometry and things that are calibrated by human understanding, engineers, understanding of the world and driving, but I suspect that's going to be a. I mean, what do I know? I'm not into self-driving I don't do self-driving cars but I know that, like Tesla is and Elon Musk is, and so their goal is to make some. They started, like everyone else, with engineering models, but I think my understanding now is that they're building more conceptual models that are based on artificial neural networks and so, rather than starting with geometry and understood things, they're just getting massive amounts of data and training it to make a model. We need a model that is at the level of high level consequences, not at the level of low level things like pixels and video. So one way you do that is by having state features that are at a more advanced level. You say, oh, this is a car rather than this is a video frame. So basically it's as simple as you need abstraction in both state and time. Abstraction in state is like saying there will be strawberries when I get to the market and abstraction in time is saying, oh, I can go to the market and then in 20 minutes I will be there, probably, and other things will be the same or related in natural ways. So we want to be able to think about I could go to the market. You also want to think, oh, I could pick up the Coke, can I could move a finger, and that will have certain consequences. All these things that we know you think are vastly different scales. Going to the market is like 20 minutes taking a new job. It might be a year deciding to study a topic. Also, it might be a period of time we think and we analyze the consequences. Like, you wanted to meet with me today and we arranged it. We set it up. It was your planning to place over weeks and some cases months, and we assembled the event of this interview by planning all that and exchanging high level messages. All that is silly to think that that is done at the level of imagining videos that we might see with our eyes or our audio signals that we might hear. So we need models that are abstract in time and state and, as a reinforcement learning person, there's a particular set of technologies that I naturally turn towards to do that. The prediction is based on multi-step prediction by temporal difference learning. The planning is done by dynamic programming, essentially value iteration, but where the steps are not low level actions but they're called options, they're high level ways of behaving that terminate. So there are things like going to the market and they will terminate when you're at the market. So at a certain conceptual level, it's clear where we want to go. To me, with abstract models in time and state, built options and features, I don't know. We did write one paper recently published in AI Journal on the notion of planning using sub-problems on the stomp progression. Stomp means subtask, option, model and planning. Put all those things together and you can do the full progression from the data stream to abstract planning, and that's what we're trying to put together. Yeah.

**Craig Smith:** 45:59

Yeah, and I sort of miss both talking about Gaia, one about that model. I mean, the input is video. It creates a representation and it plans and takes action in the representation, plans actions in the representation space. You can then decode that into video to see what it's doing, but you're not planning in the video space. So what's your ambition with this? You'll figure out the refined algorithms, the reinforcement learning algorithms. They need to be scalable. Once you have that, then you move on and start scaling them with compute and following your roadmap. Or am I simplifying it too much?

**Richard Sutton:** 46:59

We want to understand how the mind works, and then we're going to make a mind or some minds or some amount of mind, and this will be useful in all ways, in all sorts of ways, economically useful. It'll also be useful to us to extend the capabilities of our own minds. If we can understand how our minds work, we can augment them so that they can work better. Yeah, the key step is understanding, and then there would be millions of uses. I don't think it's going to be as simple as making workers sort of like slaves for us to direct. I don't think it'll be as simple as that. That maybe gives a lower bound on potential utility. Sort of our story for Etkin is we say that, well, if you suppose you could make a virtual worker, this would be enormously useful. Much of the work that we all do from day to day is it doesn't require a physical presence, it doesn't require a robot. Much of what we do is just shuffling information around. We can do most things through a video interface. So why can't we make workers that are extremely useful by playing the roles that people play? In many cases? That's sort of a lower bound on what can be done. I think much more can be done and there will be much more interesting things to be done. And then this question of what should be done. Yeah, those are rich philosophical questions and practical questions for the economy.

**Craig Smith:** 48:54

Yeah, just yeah.

**Richard Sutton:** 48:58

First step is to understand.

**Craig Smith:** 49:00

I've seen your well. And one thing on reinforcement learning and sort of supervised learning sort of took over for a while. Now it's transformer based generative AI, but during the supervised learning phase the argument was that higher knowledge is all supervised learning and the it's still supervised.

**Richard Sutton:** 49:27

It's still supervised In generative AI large language models. The training information is the next token, the next word, and that's taken as the correct action.

**Craig Smith:** 49:40

The analogy you gave me was because the analogy that's always given is that you know, a child sees an elephant, the mother says that's an elephant and the child very quickly can generalize and recognize other elements elephants. Maybe it makes a mistake and the mother corrects it and says no, that's a cow, and that was always given as an example of supervised learning. But maybe it's reinforcement learning, maybe it's the child's reward from the mother praising him for remembering the label.

**Richard Sutton:** 50:18

The point is that a child has well-developed concepts, classes concepts before and then when it's you know when its mother says that is an elephant, there's already an extensive understanding on the child's part. You know what the space is, what the objects are and this the thing that is being labeled. The label is the least interesting part of that and the child has already learned all the other most interesting parts of what it means to have animals and moving things and objects in its world. The label is the least interesting part.

**Craig Smith:** 51:07

Well, first of all, you're talking about agents that could be virtual workers. Already using reinforcement learning, people are building agents and using large language models and knowledge bases to carry out tasks knowledge-based tasks. So what you're talking about is more than linguistic tasks or knowledge-based tasks. You're talking about physical planning and physical tasks. Is that right?

**Richard Sutton:** 51:48

The key thing is having goals. If you have, for example, an assistant help you plan your day, organize your day or do tasks for you, I'm thinking it's very important that the system is able to have goals and is able to understand your goals. I think it's probably the most important part of an assistant is to understand the purposes involved. Large language models don't really understand the purposes involved. They will appear to a little bit, but the corner cases always come up. Once you spend a bit of time, you're always in a corner case. So an AI system that, after a bit, does silly things that don't respect the goals that you have or that have been given to it, that's not going to be a useful assistant. So I don't want to be critical of large language models. They're very, very useful, but it shouldn't be viewed as a criticism to say that they're also, at the same time, have rather important limitations.

**Craig Smith:** 53:01

Are you concerned at all? Are you ascribed to the threat debate?

**Richard Sutton:** 53:10

No, I think the doomers are not just wrong, I think they're blindingly biased. The bias is blinding them to what's going on. Basically, ai is a broadly applicable technology. It's not like nuclear weapons, it's not like bio weapons. It can be used for all kinds of things, and it's not the way we deal with such things is we try to use them well. There will be people that use them for bad things, and then this is just normal. Normal technology can be used by good people or bad people. The doomers are just saying oh, somehow it's bad, in the same way that nuclear weapons are bad, and they're just blinded by that metaphor, by the thinking that the AI will be out to kill them. It's just silly. The doomers don't actually give coherent reasons for what they believe, and so it's hard to argue with them. So maybe it's fair just to hold that they're biased and blind. I don't accept an argument.

**Craig Smith:** 54:35

So where you say you're maybe at stage four in the research, Carmax says 2030. It's far enough out there that maybe people won't remember in 2030 that he said 2030.

**Richard Sutton:** 54:53

2030 has been out there for a long time. It doesn't recede. It's always been 2030 for the computer power, reaching human scale quantities, but anyway, 2030 is a reasonable target for us, understanding everything that we need in order to make a real mind. Yeah, I'm good with that. You have to be ambitious. I've always said that 2030 is 25% chance of achieving a real intelligence, a real human level intelligence. 25% chance? So probably not, but it's a big enough chunk of probability that an ambitious person should work towards it and try to make it true. And it does depend upon what we do and not just the unfolding of the universe, so we should try to do that. That is a big thing. That's happening right now is the public is coming to grips with what it means for us to understand the mind and to have the ability to create minded things, and so that is a big transformation, a big change in our worldview, and so we absolutely need all kinds of people to help us become easy and have an understanding of what's happening as we achieve human level designed intelligence.

**Craig Smith:** 56:39

That's it for this week's episode. I want to thank Richard for his time. If you want to read a transcript of today's conversation, you can find one on our website. I on AI, that's EYE-ONAI. In the meantime, remember the singularity may be getting closer, but AI is already changing your world, so pay attention.