**Viren Jain:** 0:00

More importantly, we had computers that could finally do something useful with respect to this goal. They could first of all store the images and look at them on a monitor, and then, over the past 20 years, we've built increasingly sophisticated machine learning systems that can analyse the images and trace the neurons and find the synapses. Ultimately, the goal is to understand how the brain works. That's why people fund this type of work. That's really the first order of business, and you know. More specifically, the reason people want to understand how the brain works is to improve human health.

**Craig Smith:** 0:31

Hi, I'm Craig Smith and this is Eye on AI. The structure and function of the brain has long inspired artificial intelligence research. While AI has helped neuroscientists understand the brain, in this episode we glimpse into the painstaking efforts to reverse engineer thought by mapping the brain's neural pathways, a process called connectomics. Viren Jain, who leads Google's connectomics team, discusses the history of the science, from early microscopic investigations to modern large-scale electron microscopy datasets aided by AI. Viren shares his perspectives on how connectomics could inform our understanding of learning and intelligence, and talks about Google's participation in a National Institutes of Health initiative to map 10 to 15 cubic millimetres of mouse brain over the next five years. I hope you find the conversation as fascinating as I did. This episode is sponsored by Solonus, the global leader in process mining. Ai has landed and enterprises are adapting, giving customers slick experiences and the technology to deliver. The road feels long, but you're closer than you think. You see your business processes run through many systems creating data at every step. Solonus reconstructs this data to generate process intelligence, a common business language. With process intelligence, AI knows how your business flows across every department, every system and every process. With AI solutions powered by Solonus, enterprises get faster, more accurate insights, a new level of automation and a step change in productivity, performance and customer satisfaction. Process intelligence is the missing piece in the AI-enabled tech stack. Search Solonus C-E-L-O-N-I-S to find out more.

**Viren Jain:** 2:50

So, yeah, I'm Varene, I'm a research scientist at Google. I lead the connectomics team there. You know, in terms of my background, you know I got a PhD at MIT, working sort of at the intersection of computer science and neuroscience with Sebastian Sung. And you know, when I first started graduate school, you know we were working on sort of more theoretical ideas. You know, sort of how you mathematically think about how brains work. But we sort of realised that we didn't really have the data we needed to think about that or to advance that in quite the way we wanted. So then we turned to this other enterprise of really just trying to actually map out brains in detail and sort of, from the bottom up, build the data that we needed. So after graduate school I went to a place called HHMI Janelius. The Howard Hughes Medical Institute has an institute devoted to these kinds of topics, and I ran a lab there for a few years. And then, in 2013, I moved to Google where I started this connectomics team and for the past 10 years have been working on and developing the effort associated with that here at Google Research.

**Craig Smith:** 4:00

I mean, there's been a lot of work done on single cells and there's one particular worm that they've pretty much mapped. Can you give us that history before we get to Jeff Lickman and Lice? Absolutely absolutely yeah.

**Viren Jain:** 4:18

Yeah, so I mean you know the history of the field was. You know that goes back in, some say you know, even to the earliest days of neuroscience, where Cajal was, you know, about 150 years ago, looking through microscopes and standing neurons in different organisms. And you know what he noticed, and you know this is really a key insight in some ways is that neurons have different shapes. And you know these shapes are interesting because they support the general organisational principle of biological brains, which is that you have these units called neurons and they, you know, form connections with other neurons. And you know the hypothesis was that there was a lot of information, a lot of, you know, important functional information encoded in you know how the neurons actually talk to each other, you know which neurons or connections with each other ones, and so on. And so this idea kind of, you know, started to take root in neuroscience a long time ago. And then in the 70s, Sydney Brenner, who was a scientist at the MRC in London, at the laboratory for microbiology, he decided to, you know, pursue what was at that time you know, a very novel idea, which was you know what, if we just mapped out you know all of the connections in a small organism and he chose this organism called C elegans, which is a roundworm with 302 neurons, had already become popular in genomics research and was beginning to become, you know, more popular for other forms of biological research. And so they, you know, they took one of these worms, they sliced it into you know a thousand sections. They sort of image each of those sections with an electron microscope and then, interestingly, they printed them out. So they took all these sections, these, you know, basically literal films. They printed them out onto different sheets of paper and then for about you know 10 years, or maybe five to 10 years, they sat around tracing the neurons from one section to another to figure out where all the wires are going and who they're talking to, and from that they were able to, you know, reverse engineer this so-called connectome describing the chemical synaptic connections within that worm brain. So that effort has actually been quite influential. You know, people who study worm brains use that information all the time to guide their experiments. But it was such a pain, right, you know, 10 years, essentially for 302 neurons, that people you know became discouraged from trying to repeat the feat, you know, for anything larger than a worm brain, until, maybe, you know, about 20 years ago, around 2004,. People kind of revived this idea and thought you know. Now you know we can do two things. First of all, we can automate the process of sectioning and imaging tissue. In this way, you know, we can build automated platforms that will slice up tissue and image them. But, more importantly, you know, we had computers that could finally do something useful with respect to this goal. You know they could first of all store the images. You know we can look at them on a monitor. And then, over the past 20 years, you know, we built, you know, increasingly sophisticated, you know, machine learning systems that can analyse the images and trace the neurons and find the synopsis. So, you know, over the past 20 years, this field has really progressed. You know, much more rapidly. There's been a lot of progress on mapping the entire fly brain. There's, which is some work that you know our group has participated in. There's been a small amount of work done on mouse and human brain. So you mentioned Jeff Lickman earlier. You know we worked with his lab to reconstruct a cubic millimetre of human brain tissue, which was, you know, a very interesting project. On the one hand, you know, one of the largest data sets ever collected in biology. You know 1400 terabytes of data, but on the other hand it was one one millionth. So that's, you know, one over one, with six zeros in front of it, one millionth of an entire human brain. And so you know, the technical challenges in scaling this up are really quite daunting, and you know what the field is converged on is the idea that, after the fly, the next big milestone could be the mouse brain, which is about 1000 times larger than a fly, but still 1000 times larger, sorry, smaller than a human brain and is obviously, you know, a nervous system and general organism which is, you know, studied commonly in biology.

**Craig Smith:** 9:04

So for the roundworm that involved this machine, is it the same machine that Jeff Lickman uses? And was that then the same method for the fly brain that you're taking these? I don't know what the thickness of the slices of the brain is.

**Viren Jain:** 9:27

So you know, conceptually the approaches are all similar in that you know what you're doing is, you know, using some combination of cutting tissue at very thin scales. So you know anywhere between 10 to 15 nanometers, you know it's kind of like human hair thickness or smaller, and then you're imaging the tissue using some kind of electron microscope. Now the details can get quite different, though, and just to give you a flavour of this, without, you know, getting too technical, you know there's, broadly speaking, two different methods in the field, something called serial section approaches and another one called block face approaches. And in the serial section approach, so this is really like the deli slicer idea, where you take a brain, you know, you slice it up into independent sections and then you image each of those sections independently. And you know you have to be very careful because those slices are very thin and delicate. But the advantage is, you know, once you have the individual sections, you can image them with whatever kind of microscope electron microscope you want. The other approach in the field, which was, you know, used to do some of the work which I mentioned, is the block face approach. And you know, just real quick. The idea is you have like a block of tissue, you bounce electrons off the surface and then you ablate. You know you either shave off or literally vaporise the top layer so you can expose some more tissue to be imaged, and you keep alternating those two steps. Anyway, the point is there's. You know it's a real technical challenge to image a piece of brain in 3D at an animator resolution, and so there's been a lot of ingenuity put into it. You know how you can do this process reliably, at scale, with higher throughput, and that continues to be a big debate in the field. You know, just like in genomics, where you know there was a 30-year you know progression of technology that continually made sequencing genomes cheaper and more efficient. You know, in many ways we're trying to do the same thing here and try to invent, you know, the raw data acquisition techniques which could image brains at scale and more cheaply than in prior projects.

**Craig Smith:** 11:38

Yeah, and once you have the images of the slices stained, you were saying these guys I don't know if you meant literally print it out, sheets of paper and we're tracking neurons from page to page. That's all done with artificial intelligence now, isn't it?

**Viren Jain:** 12:00

That's right. So, yeah, certainly we're not printing anything out anymore, thank God. You know that would be a lot of paper, for a library even, and so computers are used in many respects. So, first of all, just to store the images and organise them and recapitulate them into a three-dimensional object. That's certainly one task that needs to be performed. Second is, you know, as I mentioned, this process of tracing neurons. So the way to think about this is you know, you probably, you know, have heard of like server closets, right? where you have like a stack of computers and in the back there's a bunch of wires connecting them, right, you know, imagine taking a black and white photo of a server closet and then trying to ask, okay, what's the wiring diagram of? You know all of how the computers connect right, the only way to do that is to, you know, find a wire and visually trace it until you end up at some other machine. Right, there's nothing. They're not colour coded, you know, there's nothing. You know there's not some number on the wire which tells you where it starts and where it goes to. And so the way to think about the brain, when we image it with electron microscopy, is like, it's like a three-dimensional server closet. And so if we want to figure out the wiring diagram, we have to literally trace out, you know, where one wire begins and where it ends. And that's quite a difficult task if you think about, you know, each of these wires is, you know, can sometimes be like tens or hundreds of nanometers thin and they span, you know, millimetres to centimetres of tissue and they're all jumbled together in like this giant mess of spaghetti. So doing that process accurately, you know, and at scale, has been non-trivial to automate but is something which you know. A number of us have put a lot of effort into designing machine learning algorithms. For, that said, you know, the algorithms are not perfect. So the final stage of all of this is actually to have humans go through and fix all the mistakes, or fix all the things that the humans, that the sorry, that computers cannot do and that remains, sort of like. Maybe the biggest bottleneck in this whole process is that, you know we still need, you know, thousands to tens of hundreds of thousands of hours of human effort to finish a given nervous system, and you know one of the big, you know, practical and research challenges in the field for us is to continuously drive that number down. So the next time we do this, we need only 10% of the human effort that we did last time, and so on and so forth.

**Craig Smith:** 14:25

Yeah, and just on that point, I was just talking to somebody about reinforcement learning and with AI feedback and having multiple AI agents sort of debating with each other to converge on a final answer. You know, to try and combat hallucinations and that sort of thing. Are there strategies like that where, if there are errors in the map, in the connectome that's been traced through machine learning, that you could have adversarial algorithms that eventually figure out? You know what the accurate path is?

**Viren Jain:** 15:19

There have been attempts to sort of combine the predictions from you know some diversity of algorithms to improve the overall performance of this tracing process. You know it has been difficult to do that in a really sort of meaningful way. Just, it tends to be the case for this task that if one algorithm is wrong, many of the others will be wrong also, but there have been attempts to do that. For now I would say the push has really been. You know, once we've had some amount of human effort applied to the hard problems or the hard cases you know, can we go back and retrain you know the initial algorithm with all that additional data to have it be more accurate in the next instance?

**Craig Smith:** 16:02

Yeah, so with the mouse brain, how far again? What percentage of the overall brain have you completed and how long did that take and all the metrics surrounding that?

**Viren Jain:** 16:22

So the mouse brain project, that's really, you know that's just getting going. So that's you know. The announcement last week was really you know, a suite of projects that the NIH has funded to really scale up mouse brain connectomics. You know what's been done there so far has also been at, let's say, the cubic millimetre scale which you know. If you think about it, a whole mouse brain is 500 cubic millimetres, that's one 500th of a mouse brain. That's sort of been achieved so far. And the goal of these new projects that the NIH is funding, of which you know we're part of one of them, is maybe to do, you know, 15 cubic millimetres or so. So you know, scale up maybe 15x and get you know, let's say you know, a few percent of the whole mouse brain at the end of the day. That said, you know that sort of the goal for the next five years is to hit that milestone and see which technologies are scaling best and you know are producing, you know, high-quality data. If that's successful, you know it's possible, the government will fund a second five-year effort which will really try and do the whole thing. So to pick a particular technology which seems to be going well and then put in the money and resources to do the complete map of the mouse brain based on that.

**Craig Smith:** 17:39

And the. What do you learn from the connectome? That's one question, because this is looking purely at the physical structure, not at the biochemistry or any of that. And then, well, why don't you answer that first?

**Viren Jain:** 18:05

Yeah Well, there's a real sense in which we're still learning what we learn from the connectome, but we've had, you know, I would say, a number of interesting examples so far. So first of all, in C elegans, where we've had the connectome for a long time, you know that data structure, so to speak. Having that information in C elegans, you know, has been transformative for studying that nervous system. So people could, for example, go in and ask okay, what are the neurons that are connected to the motor output of the system, or the sensory input, or what are the ones which you know are just sort of in the middle of all the processing? And you know, if you want to then study those particular aspects of the nervous system, you know where to begin. Or if you want to ask questions like how does the information flow, you know, accomplish some objective, some behaviour that the nervous system has, that connectome will give you a number of different clues and hypotheses which you can then use to decide your experimental strategy. Similarly, you know, I would say, the fly brain, which is a much more recent story. You know, the connectome has really been a big success there. I would claim in the sense that you know almost anybody who works on studying the fly nervous system of which there might be more people than you imagine you know basically now uses that data structure in all kinds of ways to guide their experimental strategy. So you know, you can sort of trace the information flow throughout the brain and say, okay, well, comes into these receptors. And then it goes to this part of the brain that processes, let's say, smell, information, and then it goes to the part of the brain that stores memories, then it goes to a part of the brain that decides what action the fly should make. And you know, we've really gotten to the point in the fly brain where for some specific behaviours, you know, you can literally trace out the synaptic steps that are involved in, you know, computing some piece of information or making a decision, and, you know, start to really build a detailed model and hypothesis for how the brain works. So I would say, you know, in the worm on the fly, there's now, you know, very good support for the idea that having the connectome can be extremely valuable for, you know, understanding how the system works. I think as we scale up to, you know, a mouse and one day maybe, a human brain, things will get a little bit more confusing, right? So I mean the difference. You know, one of the main differences between a worm and a fly and us is that, you know much of our brains are there to achieve, you know, learning and memory and experience driven changes in behaviour, right, you know a worm is largely born, you know, with the behaviours it's, you know, destined to accomplish. A fly, you know mostly, although it does do a little bit of learning. You know. What's interesting about, you know certainly, human brains is. You know we do a ton of learning, right, you know much of our. You know day-to-day behaviours are driven by things we learned after we were bored. So in that sense, you know how interpretable the connectome will be if you know some of these larger systems and what form these explanations will take. Of neural circuits it's a little bit less clear, but in some sense you know we were excited to find out.

**Craig Smith:** 21:25

And are there any sort of principles from now that you've got the roundworm I'm not going to try and pronounce the name of the roundworm and the fly that you that you can see principles that apply to each, maybe in the shape of neurons carrying certain kinds of information and that sort of thing.

**Viren Jain:** 21:54

Yeah, absolutely, I mean. So, you know, one of the most, I mean one of the most striking things in both, I would say, the worm and the fly, is how closely related we found structure and function to be, where the particular shapes of the neurons and you know, the connectivity that we observe just in the connectome really does seem to excuse me, you know have a close relationship with the function of those neural circuits that we observe when we go in and measure their activity and how they influence behaviour and so on. So there are parts in the million nervous systems in which we see a similarly strong relationship between structure and function. So, for example, in our retina, which is, you know, the part of the brain that's at the back of the eye and first processes light, you know, we see again a very strong relationship between sort of the shapes of neurons and they're synaptic connectivity and what their functional role is. So you know, I think that is, that is one principle which really has quite a lot of power and you know, has been, you know, proven to be true in some of these smaller organisms. Now, you know, I think, as we move to some of the higher processing centres, so to speak, in the million systems like the cortex. You know. There it seems a little bit more like a computer to a first approximation, where you know you have a bunch of you know sort of generic parts that get wired together in all different ways and the relationship between structure and function there may not be quite as obvious. And so I think you know those will be interesting. You know, cases in which you know the analysis that we do of the connective might have to be much more subtle and complicated in order to figure out what's going on.

**Craig Smith:** 23:45

Yeah, so you can see that shape and function in these lower level brains, or is that also applied to the lower brain in the human? But you can see a connection between shape and function, Are they? Is there some sort of a commonality between this kind of shape in the roundworm is analogous to this kind of shape in the fly and in their performing similar functions? I mean can, are there any? Is there any deduction that you can make Right From that?

**Viren Jain:** 24:32

yeah, so I would say the general concept which you're getting at is you know what we call cell type, right? So can we cluster, you know, all the neurons that we see in one brain, or cross brains into categories, right, and say this is one type of neuron versus another type of neuron, and based on doing that, can we then infer you know, other properties of them, like what their functional role will be in a circuit and so on? And this is a very, you know, this is a very active area of investigation for neuroscientists. So this isn't just about connectome. So there are people who are, you know, measuring gene expressions, see, you know which genetic programs are active in different brain cells, and using that to categorise cell types and so on. But in general, yes, I mean the answer is definitely that you know, there seem to be, you know, some discrete set of types of cells which you know have different genetic specification and different shapes and different physiology to go along with that, and different roles in the nervous system. You know, I think going from you know, the worm, you know, to the fly, that's a pretty big evolutionary jump actually, and it's, you know, it's not so obvious that the types we find in the fly are really that closely related, you know, to the types that we find in the worm. You know the worm is kind of an interesting case, right, because you know they've been around a long time, right. You know hundreds and millions of years at least. And you know they're small, they're 300 neurons, so they've just been hyper-optimised by evolution, you know, to do what they do efficiently. And you know it's almost kind of like a boutique analog circuit, where every part in the system is a little bit different and plays multiple roles at the same time and so on. And this is very different than if you go to, let's say, a human brain, right, which has 100 billion neurons. There's no possible way the genome could encode details of 100 billion different neurons. So what you have is a much more modular organisation, right, you have a small number of cell types, maybe thousands, which get repeated into different motifs and organised in different ways, and so you know the sort of scheme of the nervous system there is very different. So the fly is then kind of like in the intermediate, where you know there are definitely some very specific cell types. You know it's maybe not all genetically encoded, but quite a lot of it is. We see a little bit of learning and memory, so we're kind of getting a taste of both styles of computation there. But you know, coming back to your question, the identification and characterization of cell types is, you know, one of the most pressing tasks in neuroscience, because you know that's really what will let us simplify in some ways our models of these systems. You know, if we have to think of each neuron as sort of its own special thing, you know it's going to be impossible to really, you know, develop a description that's interpretable.

**Craig Smith:** 27:36

The other thing is, from what I understand. Again, I'm just a journalist Q neurons, the connections of neurons in the brain, can change over time, and even if the shapes can change over time, if I'm not mistaken. I mean so when you're slicing a brain, you're sort of capturing a moment in time. How do you track the changes in neural patterns?

**Viren Jain:** 28:15

That's right. So I mean definitely with connectomic methods we're looking at a static point in time for that particular organism. The method is highly fatal, I would say. Your brain ends up as a series of slices, so there's no more activity or changes. So it depends on the organism that you're studying. Again, for worms or flies, there certainly are dynamic elements there, but they appear to be maybe less critical to the overall story as compared to the human brain where again you can learn entirely new behaviours within a short time span and obviously encode new memories and so on. So I think studying static connectomes will let us make progress on quite a lot of aspects of the nervous system. Even in human brains it's not that things are getting completely reorganised or even mostly reorganised from one day to the next. In fact most of the structure is very likely to be quite stable over long periods of time because we have to support all of this stable physiology and all the sort of stable behaviours that we have over many decades. So I think the connectome will be extremely useful in analysing those types of issues in neuroscience, I think specifically for learning and memory. It's a bit more challenging for these methods but certainly it's not impossible. People have been looking at ways to, let's say, you take two different flies that have had two different sets of experience and you look at differences in their brain. Can you pinpoint that using methods from connectomics? But I should say neuroscience has many different experimental tools available to it. There are people who just image the activity in the brain or just look at specific connections over time and really it will be the combination of all of these different measurements and experimental tools which will be required to explain really the most complex issues in the brain, like how do we encode a memory or how do we learn things.

**Craig Smith:** 30:29

Yeah, when you're doing this you're taking cross sections. At some point of neurons You're obviously not able to slice so that you have one layer of whole neurons and then another layer of whole neurons. How much can you see into the internal structure of neurons and axons and dendrites and all of that?

**Viren Jain:** 31:02

Yeah, that's an excellent question actually, and quite a bit. So what you have with an electron microscope is really a very high resolution, unbiased instrument for probing the structure of biological tissue. So with connectomics people typically talk about okay, well, what's the list of connections between cells? But really you have much more detail that you're getting out of these data sets. So, first of all, you have the precise shape of each cell, which can often be very interesting, and then, as you mentioned, within each cell we can often see all of the organelles, so all the mitochondria, the nucleus, microtubules, all kinds of different internal structures which support the physiology of the cell. And an increasingly interesting goal in our field is to understand how those details are related to what the brain is doing overall? So if a neuron has one function versus another, does that imply that its internal organisation, the kinds of organelles it has and where they're situated, is different? Another very important structure are the synapses, which are the actual site communication between two neurons, and we can often in our pictures we can see them and we can see the individual vesicles and we can see their shape and size, which is correlated with various aspects of their physiology. So characterising all of that detail in the images is a big job and something which, again, we're applying a lot of machine learning and AI towards automating, so that in our descriptions of these circuits we don't just have this super abstract. A connects to B, but we have a lot of information about A and a lot of information about B and what their internal structures might be like.

**Craig Smith:** 32:51

Yeah, it's just making me think. When I was doing some reading about Jeff Lickman's work I don't know if it's his lab, but they produce these beautiful images of stained cross sections of the brain that you can buy and hang as wall art. Is that the resolution that you're working on, or are you working at a greater resolution than what I'm referring to?

**Viren Jain:** 33:31

So I guess I'm not exactly sure which images you might be referring to. I can say that the data we work with, the resolution, is measured in the nanometers, anywhere from 5 to 10 nanometers for each pixel. You can always go higher. So if you want to see the internal structure of an individual protein, then you have to go to a single nanometer or higher resolution. But if you want to see how a cello is organised, you can see quite a bit at a few nanometer resolution. And that is what our entire data sets comprise. And the reason they're so high resolution is that the machinery that the brain uses to establish these connections is the synapses and the individual axons. They often become very, very small at that scale of nanometers. So if we want to actually resolve all these connections, it turns out we need to both image at very high resolution in order to see those connections, but also over very large fields of view in order to see where they go. And from a technical point of view, that's really what makes this field very challenging. We have lots of methods which can image large brains, if that's what you want to do, and we have lots of methods that can image at very high resolution. But here. We have to do both at the same time in order to make sense of things.

**Craig Smith:** 34:57

And does your work involve? You're doing the mapping, but are you also making hypotheses about the function of different shapes or different structures that you're seeing within the, or the different connections for how? Not only how information flows, but what kind of information flows through different neurons? I mean, for example, visual information or some other sensory information.

**Viren Jain:** 35:45

Right. So you know, I mean, in terms of the field as a whole, those are definitely, you know, critical topics which people study using this data. You know, at Google Research, you know we've chosen to specialise on the parts of the process where we can sort of add the most value, and that's really. You know we don't collect the raw data. You know we work with people like Jeff or HHMI or Max Planck, so they're collecting these huge data sets and then, you know, we sort of sit in the middle, we sort of take that data, we turn it into a form that's, you know, annotated and traced and then is usable by biologists downstream who might be interested in all sorts of questions. So you know, we don't really see our role, as you know, necessarily, you know solving the biological questions or inferences, but really as just taking the raw data, applying a ton of computational tools and techniques and machine learning algorithms to turn it into a form where then a biologist you know who is interested in some particular question could use it. And so you know, to do that, you know we've developed ML methods to solve that tracing problem I mentioned, or you know, as you were just referring to, to solve this problem of cell types. You know. I just give you a small fragment of a cell. Can I tell you, you know what type of cell that is, because then you'll know a lot more about what to expect its function to be. And that's a problem which you know we've worked on and have a paper coming about, coming on about in a few weeks. So that's really where we see, as you know, the area which we try to focus on.

**Craig Smith:** 37:20

And what is Google's interest in this? Because is it? Is this a service that you're providing to the research community, or are you building a product? You know, whether it be data sets or techniques, algorithms that then can be applied elsewhere, because it can be applied to the research community.

**Viren Jain:** 37:48

Yeah, yeah, absolutely. I mean. You know this is primarily a research endeavour. So you know we're not charging our collaborators. In fact, you know, with this new project with the NIH, you know we are working with a bunch of people who you know they're getting money from the government, but we are not. You know we are contributing our, our, our, our work and our, our resources, you know, pro bono, towards this goal. So you know, I think there's a, you know, to see how computer science can actually have a useful impact on some of the fundamental problems in. You know the natural sciences and you know this was one area where there were these huge data sets and these machine learning problems and it just seemed like a good fit for the kinds of things which you know Google has historically excelled at. You know. You think about, you know, mapping the web, mapping the earth. You know why not jump in and help map the brain. That said, you know there have been, you know, things that our team and collaborators have achieved which have been useful in the company more broadly. So, for example, you know one of the most basic things we had to build was infrastructure to store and manage all of this huge data right. So, you know, we get literally millions of images which then need to be, you know, recapitulated into a three-dimensional ported system. We need to annotate them, visualise them, process them. To do this we built a system called TensorFlow and it turns out that system, you know, has been useful for other teams at Google. So, for example, the teams that are training these huge large language models or other machine learning systems ended up using that to store all the parameters of those models. So you know, there is, you know this general, you know research intuition that if you choose a really hard problem and make progress, you'll end up inventing things that are more broadly useful. And in computer science that's especially likely because you know the problems end up being very general, you know. So if you pick a really hard you know machine learning problem or data problem, you know you'll kind of have to invent things that are more broadly applicable.

**Craig Smith:** 39:51

Yeah, and then so that's how computer science or AI is aiding in this research or this discovery. And there's also does this kind of work give you any insight into how the brain learns? And the reason I'm asking is I talk periodically to Jeff Hinton, whose primary goal is to understand how the brain works, not to build, you know, big, profitable AI engines, and so his frustration has been that back propagation of, after you know, decades of research is the conclusion is that that does not use. That kind of an algorithm does not exist in the brain, and he, more recently, has worked on something called the forward forward algorithm, which he thinks could possibly work in the brain, where information is a past back and forth, but not back along the same pathways. And is that kind of thing something that you could find evidence for, or is? Are we no longer talking about the physical structures and more about well, you would be, ultimately, because you're talking about, you know, electrical impulses and changing weights, and from now on, this is a great question, greg.

**Viren Jain:** 41:46

So you know we have a project actually where we're looking at this. So I'll tell you a little bit about the project and the species. The organism which we're studying in order to shed light on this is something called a zebra finch. So it's a bird, and the interesting thing about these birds is that they learn their song from their tutor, typically their parents, but some other bird when they're young. They're born not knowing a song and then, over a period of weeks, they learn a very specific song, and neuroscience has been very curious about this process because you know it's sort of a stripped down version of the learning problem, right? I mean, how did they go from not knowing a song to knowing a very specific song in a matter of weeks? And you know it's interesting precisely for the reason you mentioned, because you know, in general, neuroscientists do not think back. Propagation is sort of a plausible explanation for how such a learning process might work, and so there have been a number of, you know, very specific theories put forth about how, in this bird, this might work. And you know, behaviorally it's interesting because the bird starts out babbling right, making a song that's incorrect, and then they slowly and slowly get better right over, you know, a period of days to weeks. They keep tweaking the song and each time it's a little bit more like the song that they're supposed to make. And so people have identified a bunch of circuitry in the song brain, the songbird brain, which seems to be associated with this process. And what we've done now in a preprint and a project that is still ongoing is we mapped out the circuitry and in one part of that part of the bird brain and indeed we find evidence for a very specific type of learning algorithm, something called node perturbation, which is an alternative to back propagation. And you know, this is basically a mathematical alternative to the type of gradient learning which back propagation achieves, which was proposed, you know, 40 years ago. And you know, we found, you know, some very specific types of connections and patterns of connections which would support this type of learning algorithm versus others. And you know, this is just the beginning of this type of work. You know it's not easy and, like the inferences are very subtle at this stage, but we're sort of excited to take this further because, as you mentioned, I mean, this is one of the key problems in neuroscience. Right, you know brains seem to be, you know, doing all of this learning and adaptation. But if back propagation is not the answer, what is? And you know how to do it efficiently? You know, probably you know. One of the most fascinating things about, you know, biological intelligence, you know, is, for example, how energy efficient it is. Where you know a human brain uses, basically you know, a light bulb's worth of energy at any given moment, whereas you know, you know, big machine learning systems can use much more than that. So you know. Another big question which you know we're interested in is, you know, regardless of sort of, you know, ultimate performance or accuracy. You know, how is it that biological systems manage to do it with so much less energy and you know so much less computation than the artificial systems we have?

**Craig Smith:** 45:06

And so much less data also. Yeah, that's right, I had Rich Sutton on the podcast a couple of years ago. I mean it made me think. When you're talking about Finch, it is that he is learning this song and it gradually gets better. It makes me think of his temporal difference learning algorithm.

**Viren Jain:** 45:39

Yeah, absolutely. I mean, it's an example of a reinforcement learning process.

**Craig Smith:** 45:42

Yeah Right, yeah, and that algorithm, from what I understand, has been widely recognized as the algorithm functioning in the lower brain, or at least parts of the lower brain. What I mean is can you make that connection between what's happening, the algorithms at work in the lower brain of the human, and what's happening in these less-evolved animal brains which don't have a big cerebral cortex driving them?

**Viren Jain:** 46:32

Well, I would say all of this is very much a work in progress. What I would say is there's a difference between some of the different problems that are being addressed here, so with temporal difference learning and the kinds of things which that style of reinforcement learning has studied. Often that's about figuring out what's the right set of actions, right To maximise some reward or minimise some pain, or something like that, where you have an environment in which there's many things of different things you could do, what's the right next thing to do, and so on, and you get feedback only once in a while, let's say, and that's the big learning. Problem is you have to figure out what to do even though you're only getting feedback, and maybe the feedback is very vague about what the policy really should be, as the term is called, whereas with the zebra finch and the behaviour we're studying there, it's a little bit different, because you have this very precise motor activity which has to generate this complicated song, right? So the bird in some sense knows what it has to do. The question is, how does it set all these different parameters of very precise motor timing and muscle activations in order to reproduce that acoustic signature of the song it's trying to resemble and that's certainly a type of reinforcement learning problem, but it's a little bit different in the sense that it's not so much an action space, it's sort of this parameter space of all these different muscles and so forth, and so the characteristics of the learning problem become a little bit different. So that certainly is something which human brains also have to do. We have to learn how to walk and talk to ourselves and control our own bodies and so forth. But I would say, so far it seems like the particular sort of style of learning that you might see play out might be a little bit different than something like temporal difference learning, where you have sort of this more higher level planning and feedback to deal with.

**Craig Smith:** 48:37

Yeah, where is this going? So the NIH has come through with this money. You guys are working on how many labs across the country? Or are there other countries that are doing similar research? And yeah, how is this progressing? Is it progressing quickly or is the work so I don't want to say tedious, but so incremental that it's going to take a long time before you have the data on which to make conclusions?

**Viren Jain:** 49:18

Yeah, I mean, from my point of view it seems to be progressing relatively quickly at this point. I think up until the results from the fly brain, there was still considerable scepticism about whether this was worth doing, which is a natural thing in science. There's limited resources and people and money, and so we have to make bets about what's actually going to be useful. And I think up until we had the results, the community was kind of mixed. Some people thought it would be worth doing and some people really thought it was a waste of time. I would say, since the fly brain results, there's much more consensus that there's real value in this approach and we really want to have this data not just for the fly brain but for a mouse brain and maybe one day for a human brain as well. So I think now that that momentum, that consensus, is a little bit more in place, there's much stronger forces at work trying to propel this forward. So obviously I just mentioned the NIH, but there's even private foundations that are now funding work in connectomics. There's other countries. The Max Planck Institute in Germany has been robustly supporting this. The Howard Hughes Medical Institute was responsible for a lot of the fly funding. The Wellcome Foundation just wrote a whole white paper analysing the prospects for a mouse connectome. So the Wellcome Foundation is basically the largest actual nonprofit biomedical charity in the world and they commissioned basically a whole report analysing what it would take to do a whole mouse brain connectome. And the thing about progress is that it's often geometric or exponential. So if you go back to the human genome project in the 90s, this was again a big government effort in the US to sequence a whole human genome. The first seven or eight years of the project they achieved two to 3% of the whole genome. But what happened was during that time there was so much progress in figuring out how to improve the methods and how to do it more efficiently that they finished the rest of it in the last few years. And that's the really powerful thing about these technological development projects, where you keep improving the methods at the same time as you're doing the science and then one day all of a sudden you can do something 10 times faster than you could a few years ago and that completely changes how people think about things. So we're sort of in the early phases of that expansion and that acceleration for brain mapping. But I would be surprised if it didn't play out. Genomics is really a very high bar. They basically even exceeded Moore's law for how quickly things got better. So I don't know if we're going to do anything quite as aggressive. But even if it's a more modest rate of geometric progress, the answer would still be that within years to a decade or more, we would be in a completely different and transformative position for performing this kind of mapping to the extent we want to. That said, I think for all the human brain, which is arguably the logical conclusion of all this, that's probably still some decades away.

**Craig Smith:** 53:00

That's a million fold larger than what we can do, right now and the technology that you're working with to gather data is still this slicing, whether it's the block ablation or the salon.

**Viren Jain:** 53:20

That's right, I mean it's some combination of slicing and electron microscopy. That said, people are really aggressively looking at alternatives. People are even using those synchrotrons, those large circular buildings, for accelerating light, and using x-rays to image tissue. There's people working on various fancy methods of using photons instead of electrons to image tissue. So this is an area of biological method development that's extremely active and there's a lot of incentive to make things better and cheaper and more efficient, and people are extremely clever in coming up with ideas to do so.

**Craig Smith:** 54:07

So I wouldn't be surprised if, five or ten years from now, we're acquiring our data in a substantially different way than we are now, yeah, and I was thinking of Neuralink, which you know manages to get these electrodes threaded through brain tissue without damaging blood vessels or with minimal damage to neurons. Where I mean is there is Neuralink involved Because, there you're, you'd be able to gather data from living brain cells without killing them in order to see them.

**Viren Jain:** 54:51

Well, I mean, one can debate exactly what Neuralink is or is not killing along the way. But you know, you know that's a very different endeavour, I would say. So they're not really trying to map out the structure of the brain, they're trying to get something into the brain so that they can, you know, both monitor and stimulate electrical activity.

**Craig Smith:** 55:10

Yeah, but could that? I mean, is there another way that you could gather data from, from the connectome?

**Viren Jain:** 55:22

Well, you would certainly be getting information about the brain right. You would be getting, you know, recordings of some neurons or some fraction of neurons in a brain, but it would be a very different type of data than you know. Mapping out the structural connectivity, yeah, there are other methods for recording activity. So people, for example, have invented something called a light sheet microscope where you can put a whole baby fish into it and image all of them. You can. You can take a movie, basically, that describes the activity of every cell in that brain over time. And you know those are very exciting data sets. But scaling that through a human is, you know, extremely improbable, at this point at least.

**Craig Smith:** 56:05

And what's the, the, the, the end ambition is that eventually you would map this mouse brain, you'd be able to recreate it with artificial neurons in a computer, you know, with programming, and then run different algorithms. To see what I mean? Yeah, I mean that in my mind that's where this would go. If you have this computer program that is basically an analog of the mouse brain and then you're playing with different algorithms to see how the thing should be working or could work? Or am I in science fiction land?

**Viren Jain:** 57:02

Well, no, I don't think you're in science fiction land, I mean, I think, look, I mean, ultimately, the goal is to understand how the brain works. Right, I mean that's why people fund this type of work. I mean, that's, that's really, that's really the first order of business. And you know, you know, more specifically, the reason people want to understand how the brain works is to, you know, improve human health. Right, I mean, that's why the NIH funds this work, that's why HHMI and Max Planck fund this work. Ultimately, the goal of all of this is to improve medicine, which you know is quite appropriate given the public investments in this space. But in order to do that, we need to figure out how the system works and what goes wrong when it doesn't. And for the brain that's enormously difficult, right, you know, if you think about the kinds of treatments that we have for, you know, mental diseases or neurodegenerative diseases and so on, it's quite abysmal. Frankly, you know, like not a lot has changed over the past few decades. You know the drugs that we have to treat. You know, very common mental health problems or more complex ones like Parkinson's and so on, are not great. And you know there's, there's really an enormous need right for, for better options. So the main goal of all of this is to aid in that endeavour. But that's super high level, I think you know. Once you get a level below that, you know the question is you know what are the different tactics that you might take in order to make progress in that question? And indeed, you know one approach might be okay. Let's have computational systems that can reproduce aspects of the brain so that we can study the brain more efficiently. You know, if you think about, you know something like what's happened with proteins recently, where we can computationally go from a sequence of amino acids to the 3D structure, you know that dramatically accelerates the kinds of things you can explore and the number of options you can, you know, investigate. Similarly, if we had, you know, computational models of the brain or parts of the brain that were, you know, realistic, you know we can much more quickly explore. You know different options for manipulating the nervous system to treat some of these issues or figuring out what the basic parameters and principles are. So I would say those are the main goals you know, and we have a lot of work to do before we sort of realise them.

**Craig Smith:** 59:27

Yeah, but wouldn't there also be, wouldn't this inform artificial intelligence research? I mean, there could be architecture or, as you say. I mean the brain operates on very low wattage and very little data compared to, you know, a large language model, for example. Couldn't this, wouldn't there be insights to building AI models that could perform in ways similar to the brain?

**Viren Jain:** 1:00:08

So it's possible. It's just a question of you know who makes you know, faster progress in some ways, right, and you know, as far as I could tell my colleagues, you know, on the AI side of things are, you know, you know things are moving along very quickly there and you know the question is you know, to what extent do they really, you know, need insights from biological systems to keep making progress? And there's no. You know it's hard to answer that question definitively, right? You know you have different intuitions and certainly, if you think about it you know how they got to the present day. You know where these systems came from? You know a lot of that came from folks like Jeff Hinton, you know, paying attention to basic principles about how the brain works, right. So this idea that you know artificial intelligence systems should be made out of deep learning systems, right. That's not an obvious choice to make and in some sense, you know, for a long time people had other ideas, but you know it was folks like Jeff Hinton, who were inspired by very basic aspects of brain organisation, that kept pushing on that approach and that's really what got us to where we are today. Now, as to whether you know the next 20 years of progress will need some additional insight from biology. You know it's hard to know. You know if they're going to run into some sort of wall and, you know, run out of ideas, then maybe, or maybe at this point you know, they're just going to keep engineering those systems to be better and better and more efficient. You know, maybe at that point, you know, biology is kind of irrelevant. Who knows?

**Craig Smith:** 1:01:37

Hi, this episode is sponsored by Salonis, the global leader in process mining. Ai has landed and enterprises are adapting, giving customers slick experiences and the technology to deliver. The road feels long, but you're closer than you think. You see your business processes run through many systems, creating data at every step. Salonis reconstructs this data to generate process intelligence, a common business language With process intelligence. Ai knows how your business flows across every department, every system and every process. With AI solutions powered by Salonis, enterprises get faster, more accurate insights, a new level of automation and a step change in productivity, performance and customer satisfaction. Process intelligence is the missing piece in the AI enabled tech stack. Search Salonis C-E-L-O-N-I-S to find out more. That's it for this episode. I want to thank Ren for his time. If you want to read a transcript of this conversation, you can find one, as always, on our website, ionai. That's E-Y-E-O-NAI. And remember the singularity may not be near, but AI is changing our world, so pay attention.