EPISODE 91

[INTRODUCTION]

[0:00:10.4] MC: Hello and welcome to another episode of TWIML talk. The podcast where I interview interesting people doing interesting things and machine learning and artificial intelligence. I'm your host Sam Charrington.

This week on the podcast, we're featuring a series of conversations from the NIPS conference in long beach California. This was my first time at NIPS and I had a great time there. I attended a bunch of talks and of course learned a ton, I organized an impromptu round table on building AI products and I met a bunch of wonderful people including some former TWIML talk guests.

I'll be sharing a bit more about my experiences at NIPS via my newsletter which you should take a second right now to subscribe to at twimlai.com/newsletter. This week, through the end of the year, we're running a special listener appreciation contest to celebrate hitting one million listens on the podcast and to thank you all for being so awesome.

Tweet to us using the #twiml1mil to enter. Everyone who enters is a winner and we're giving away a bunch of cool TWIML swag and other mystery prizes. If you're not on twitter or want more ways to enter, visit twimlai.com/twiml1mil for the full rundown.

Before we dive in, I'd like to thank our friends over at Intel Nirvana for their sponsorship of this podcast and our NIPS series. While intel was very active at NIPS with a bunch of workshops demonstrations and poster sessions, their big news this time was the first public viewing of the Intel Nirvana neural network processor or NNP.

The goal of the NNP architecture is to provide the flexibility needed to support deep learning primitives while making the core hardware components as efficient as possible. Giving neural network designers powerful tools for solving larger and more difficult problems while minimizing data movement and maximizing data reuse.

To learn more about Intel's AI products group and the Intel Nirvana NNP, visit intelnirvana.com. This time around I'm joined by Matthew Crosby, a researcher at imperial college, London. Working on The Kinds of Intelligence Project. Matthew joined me after the NIPS Symposium of the same name, an event that brought researchers from a variety of disciplines together towards three aims.

A broader perspective on the possible types of intelligence beyond human intelligence, better measurements of intelligence and a more purposeful analysis of where progress should be made in AI to best benefit society.

Matthew's research explores intelligence from a philosophical perspective. Exploring ideas like predictive processing and controlled hallucination and how these theories of intelligence impact the way we approach creating artificial intelligence.

This was a very interesting conversation and one that I'm sure you'll get a kick out of. Now, on to the show.

[INTERVIEW]

[0:03:22.9] SC: All right everyone, I am here in Long beach California at the NIPS conference and I have the pleasure of being joined by Matthew Crosby. Research associate at Imperial College London. Matthew, welcome to the podcast.

[0:03:35.9] MC: Thank you very much. Why don't we get started by having you tell us a little bit about your background and how you got involved in AI?

[0:03:43.0] SC: I was always very interested in intelligence as a concept. I had a fairly mathematical sort of up bringing and start. My first thought was, I'm going to explore through mathematics and computer science.

[0:03:57.6] MC: Okay.

[0:03:58.4] SC: I worked for a while in high level planning in AI and over the course of the time working on that stuff, I sort of got a bit disillusioned with the fact that AI was less about intelligence than I was hoping it would be.

Because for me, intelligence or spoke in some sense to human like intelligence. To this ability to sort of experience the world and plan in the world that's why I was working in planning to form ideas about what you're going to do and represent this like picture of the world in front of you. That was fundamental part of intelligence.

[0:04:31.4] MC: We're not quite there yet.

[0:04:32.2] SC: Well, it wasn't being covered in the part of the field that I was working on. Then, as I was sort of becoming a bit disillusioned with that, I found these theories in philosophy which I'd been dabbling in which suddenly spoke to what I felt my experience of the world was like and what intelligence actually was and it was the first time when I thought, no, people are actually explaining this at a level where it's making sense to me and it's progressing us further towards a better understanding.

I just completely transitioned, went back and relearned philosophy so that I could understand these theories and philosophy of mine better. While I was doing that, we were seeing all these advances now in neural networks and in deep learning and machine learning where we are actually approaching something that looks like intelligence in a way that we can talk about – especially philosophically, that's interesting.

Makes sense and does speak to what I was interested in, in terms of intelligence. Now after making this transition to philosophy, I'm back in a sort of harsher role in between the two where I'm looking at intelligence in AI but for a more philosophical perspective.

[0:05:35.2] SC: How does that translate into a research path for you?

[0:05:38.7] MC: Well, one of the problems with that is that's obviously a giant research project and – the project I'm on right now is The Kinds of Intelligence Project which is a sub project at the Leather Home center for the future of intelligence.

[0:05:51.4] SC: Okay.

[0:05:50.8] MC: This is a large - 10 million pound project in the UK mainly based in Cambridge where the idea is to take people from multiple disciplines and look at the future of intelligence and the path we're on towards the future intelligence and take expertise from AI but also from philosophy, political science, lecture and all the broad range of subjects that could have something to say about intelligence. Use it to shape research in the future, shape policy decisions and make sure that we are moving towards a future that's best for everyone in terms of intelligence.

[0:06:26.9] SC: The Kinds of Intelligence is also the name of a symposium that you co-organized here at NIPS. Tell us a little bit about that and the goals for it?

[0:06:34.6] MC: The goals of The Kinds of Intelligence Project are to map the space of intelligence so that we can better understand the intelligence landscape. The kinds of intelligence symposium at NIPS. The idea was to bring a lot of people who were at the forefront of different types of intelligence together into the same place and so we can have this conversation and look at the way the landscape is shaped both from plant cognition to animal cognition, to child development which are all very important parts of understanding intelligence to obviously being at NIPS, AI and machine learning and the type of intelligence we can now create with – we had people talking like David Hassabis from DeepMind he was talking about Alpha Zero which is now solving, Go and Chess and Shogi at levels, way beyond human intelligence and from people talking about plant cognition and the way, if you drop certain plants from a height to over time, they can learn a response to curl up in protection from falling on the floor. You drop them enough times, eventually they learn to anticipate what is going to happen and curl up ahead of time.

[0:07:34.6] SC: Wow.

[0:07:34.7] MC: This is a form of learning that doesn't have any neurons involved and it's very alien to the type of learning that we generally think about.

[0:07:40.7] SC: Interesting. Your personal kind of slice through this is from a philosophical perspective and you mentioned some kind of a body of work or research within philosophy that you stumbled across. What is that?

[0:07:56.7] MC: The general time for this is predictive processing. It's the idea that when we're taking sentry data from the world, we have this huge jumble of messy information coming in to the system, right? We have 130 million or so photo-receptors in each eye all transducing electromagnetic information.

Somehow the brain has to make sense of that and understand it and at some point, it understands it in a way that we sort of experience the world and that's how that happened. The old, sort of very old view was that this information comes into the system and in the very bottom of where it gets pieced together, more and more complicated as it goes up through the system and eventually you get ideas such as tables and chairs and you know, the kind of objects that we feel like that we see.

Predictive processing idea sort of turns that on its head and says, we're not in the process of sort of taking these components of information and putting them together. We're actively trying to work out what we're going to experience. We're predicting the incoming sensory information and actively doing so, we're always trying to work out what's going to happen next in the world.

By turning it that way around and looking at how we could actively predict, we see that our experience of the world takes the form of what is called a controlled hallucination and the phrase is becoming much more popular nowadays. The idea is in hallucination, you're just making stuff up, right? Maybe your brain is making stuff up that's not really there and that's what you see, right? In a controlled hallucination, you're making stuff up just as you are in a real hallucination but it's controlled by the actual sensory data itself.

There's no real difference from me seeing this table in front of me right now to hallucinating one except for the fact that there's a ground truth from the sensory data that is binding it together so that hopefully, when I see a table there, from your perspective, you're also seeing a similar table.

[0:09:44.0] SC: What are the implications of seeing cognition as this controlled hallucination process?

[0:09:52.2] MC: A huge number of implications from this and I think that's one of the beauties of the theory and probably one of the potential downfalls of the theory too is that it can apply at so many different levels across the brain and it also – in relation to machine learning, we're seeing obviously a huge focus on predictive algorithms and on generative models which are generating predictions about the world, all the century data over there being input.

We can think of this as a very low level – in the retina or the back of the eye, we're doing what is called predictive coding which is, whenever I get – say, a particular rod cell is hit by electromagnetic radiation, it has the amount of information of the intensity of the light that it can transfer up further in the brain.

[0:10:34.2] SC: Right.

[0:10:34.7] MC: That could be a large number of different values at this take. If instead of transferring that value, I look at what I would expect just by looking locally at all the values of the rod sales or the cells around me. I can take the average of that and see how much that particular cell is different from that average. Then you'll get a much smaller number which increases the bandwidth that you won't – decrease in the bandwidth that you need to use this in the same amount of information.

[0:11:01.8] SC: Are you describing a theory of what is actually happening physiologically or are you describing a modeling approach or?

[0:11:10.8] MC: This is –

[0:11:11.6] SC: It's sounds like nearest neighbor happening in the eyes or something.

[0:11:14.7] MC: A bit like that, yes. But – I was starting at the point where this is actually yes, we know this is happening at a neurophysiological level and I was going to move on to the complete other side where this approach can be applied to beliefs and desires.

Where our beliefs and desires are baited in a similar sort of predictive format. Before I move on, one thing with this predictive coding approach at the retina is if you look at it, it is very much like convolutions that we're seeing in machine learning and the way they work is a very sort of similar approach.

They're obviously a bit more focused on variance individual field and how we could apply the same math but different locations which is another area of this. They could also be used to do a similar approach to predicting local variations and only transmitting information about that variation as supposed to the raw data itself.

That's one level of it but at the other level, we have the idea that our beliefs are updated in a similar way and our high level understanding of the world is based on these predictions that we're making and comparing to the data coming in and when it's wrong, we have two choices, we either update our prediction and therefore change our model of the world and see the world differently. Or, we act in the world and move around the world and that make our prediction that was wrong, turn into a prediction that was correct. For example, if I predict that table is off to my right, there's two ways that I can make that true.

Well, I can turn – I can change my prediction, I can be, no, it's wrong, it's to my left and then update it that way or I could move my head and that would make the - another way of making the same prediction right. I think this is a way that we could bring actions and interacting with the world into our understanding especially in machine learning and robotics of how we can incorporate this sort of predictive approach to the whole brain or the whole way of modeling the world into a system that's not just understanding the world but also acting in it and performing tasks.Therefore, being intelligent.

[0:13:14.1] SC: is your research kind of approaching this from a theoretical perspective exclusively or is there an experimental element or applied element as well?

[0:13:25.3] MC: I have been running experiments with predictive coding style in neural network which in a few, coming out recently based on the structure where you have a hierarchical generative network and each layer of the hierarchy is just trying to predict anything that the lower layers have so far failed to predict.

The first layer tries to predict the world but it’s not strong enough by itself to fully model everything so then, what it can’t predict.

[0:13:51.4] SC: Can you give me an example of kind of the experiment? What it’s specifically is it trying to predict?

[0:13:57.6] MC: There’s been work on this in experiments from video data from images, for videos on top of a car, moving around, driving around, you give it the first nine frames and ask it to predict the tenth.

Then, you can get fairly good results at predicting how the road is going to have moved and the things that will have come into view. I’ve been experimenting with this in maze like three dimensional domains, working on the raw pixel input and predicting how the maze is going to update or how the pixel’s going to update as I move around this maze.

[0:14:27.4] SC: Okay. I spoke with a researcher working on something very similar like she was looking at it from a perspective of like embodied computer vision. Not just fixed frames but you know, fixed frames plus the ability to move the orientation of the view port, if you will.

One of the sets of experiments or use cases was this kind of center mounted on a car that was changing direction and trying to do the prediction, things like that.

[0:14:56.9] MC: Yes.

[0:14:57.4] SC: Interesting. You’re got this scenario with the video case, you got the scenario where the – you’ve got the camera on the vehicle and you’re trying to predict the head. How does that then tie into this hierarchical structure that you were describing?

[0:15:13.3] MC: Well, the idea of the hierarchy is that you’ll have the first layer of your network would be, not have enough potential space inside it to fully predict everything that’s going on, there’s just not enough space to represent the full function of the mapping of the change, of the inputs over time.

Then, anything that it can predict, you're not interested in that anymore because that’s sort of done. That gets sort of finished at that layer of the system. Anything that doesn’t get predicted, this is called the prediction error, will get sent up to the next layer of the system. This is the – if everything’s working as intended, this is the parts that are a bit more complex than could be predicted just easily, like you know, something, this pixel always stays the same so I’m just going to predict it stays the same but then something slowly moving across the visual fields or the input space in this video might be impossible to predict to that lower level but at a higher level, once you’ve removed the easy stuff.

And you’ve got access to more machinery because you’re layers deep into the system, you might be able to predict that and the idea is that as you move up the hierarchy, you’ll get more high level representations of elements of the world that are changing.

The low level would be very simple stuff, the high level could potentially be things changing very slowly over time or changing, more modeling like being true to physics, going on behind the domain that you might expect, this object’s hitting something so now it’s going to change direction.

The lower level might think, it’s just going to continue going on in the same direction. Then when it gets that wrong, the high level which hopefully has representation of the physics will be able to correct for that and say no, this is how it’s going to change.

[0:16:51.0] SC: What you're describing sounds like what I think of a maybe a very deep convolutional net, right? That’s going to figure out different things at different levels. How is it different or are you doing things to kind of force it to learn certain aspects and certain layers or –

[0:17:08.3] MC: Yes, I think the beauty of the research is that, on one level, you could think of it as just as a very deep convolutional net. The one thing that is different is you’re focusing on this particular prediction paradigm which has its own nice properties to it. For example, if I give you a noisy input, if the video is full of noise and I’m trying to predict the next frame in a noisy video.

If the noise is unbiased, so that over time, the average of the noise is zero, then the prediction automatically filters that out at the very bottom layer of the structure. That doesn't get passed up to the higher level. They see less noisy input. You get powerful results just from moving to this prediction paradigm. Also, the second thing that is different is, the main propagation from layer to layer is the errors of the stuff that you can't predict as supposed to just a more sort of complex combination of everything you've got so far.

[0:18:03.3] SC: Does that result in a network that is – that has kind of fundamental differences and properties than what you might see in a typical CNN, like in terms of the density of the weights or the way the layers are interconnected with one another?

[0:18:20.7] MC: Yeah, I think so far, I mean, the research is fairly early here and there will be – I don't really know the answer to that question. The answer is going to be yes, but I can tell you in good details exactly how this approach differs from the others.

I think the more you incorporate from different areas in machine learning techniques, the more you're going to find that this approach looks like all the others.

[0:18:42.2] SC: Right. The layers are trained end to end, you're not training individual layers separately.

[0:18:48.1] MC: You can train the individual layers separately because each layer's input is just the – from the local.

[0:18:54.6] SC: Okay. When you acknowledge this is early, have you had any preliminary results from this line of research?

[0:19:03.6] MC: The preliminary results, the idea works and the suggestion there is that if this is a good philosophical theory of how the brain might be working and then when we do implement, actually implement it in our networks, we see positive results. That that's just a sort of a backup result to say, yeah, maybe the philosophical theory is on to something, right?

Maybe this is a good idea. It can predict future frames with high accuracy, you know, just when moving around the domain but it's still early days to say, how well that will be when we move it into reinforcement learning type experiments where we can compare to state of the art and see how that – learning that kind of representation how.

[0:19:44.4] SC: Yeah, the impression that I get from the conversations I've had recently on related topics is that our understanding of the brand and the neurophysiology and our understanding of the machine learning are – where kind of one leapfrogs the other and then feeds back you know, learning to the other and it's kind of this iterative process, is that your sense as well or is one area like you know, far ahead of the other and you know, for example, we understand the brain a lot more than we do the machine learning side and machine learning continues to pull from that or the other way around.

[0:20:20.9] MC: I think we understand them in very different ways but there are a lot of people that are sort of, on the cusp between the two disciplines that are grabbing stuff from one and pulling it into the other and grabbing stuff in the other direction. We've seen that work. Convolutions as I mentioned before across these two disciplines and they work on both sides of the spectrum.

[0:20:40.8] SC: Which side did they come from? Do you know?

[0:20:43.6] MC: I think it depends on your viewpoint. I was at this symposium, Gary Marcus was saying that perhaps the Anglican wasn't aware of all the predictive coding type work in the retina and how convolutions might be applied in low level visual systems.

[0:20:57.0] SC: Okay.

[0:20:57.1] MC: He was just trying stuff and found something that works that happens to be very much related.

[0:21:01.7] SC: Okay.

[0:21:02.6] MC: Yeah, I think that depends who you ask.

[0:21:06.5] SC: Interesting. You also mentioned earlier in kind of the conversation of philosophy, theory of mind, more broadly. Can you describe that and how that fits into all of this?

[0:21:18.5] MC: Yeah. At first my intuitions about why intelligence is interesting are that it involves introspection and thoughts and the ability to reason about the world in the way we do which is in a sense, it's sort of symbolic process, right? We construct sentences, we have language, that's sort of very key component.

We construct sentences in our heads and we understand things through those sentences sometimes. It's really interesting to see how modern work, it's starting to look at representations of the world that in machine learning, where we can answer questions in natural languages applied to images for example, we've seen relation that's where you take in an image, containing a few objects of different sizes and different locations and you answer through a natural language question such as, "Is the red job object to the left of the yellow object? That kind of question.

The way they work is to try and create a representation inside the network that understands this relational kind of information which is moving towards, in my opinion, moving towards a more symbolic or at least potential for a symbolic understanding of the world inside the standard machine learning algorithm.

[0:22:30.4] SC: My reaction to that from very little kind of reading in linguistics is that part of that is not, the idea of thinking in sentences is not universally accepted, is that right? You know, a lot of ways, thought as more abstract than sentences, there were experiments about – trying to remember the – you know, there is this line of thinking around whether the degree to which language impacts thought and you know, I think there's kind of this popular belief that people think in their languages but it's also been disproven in a lot of ways.

[0:23:08.4] MC: Yeah, I don't know too much about that area to be honest. I do think that some kind of – not necessarily language type processing but symbolic level processing or processing where we understand objects as entities that persist over time that we can therefore then label is going to be necessary as we move towards general types of intelligence. Especially if we end up on this track where it seems, at least there's a lot of key players in the field right now pulling towards human-like general intelligence. That that's going to be a key component and I think as we move towards it in AI machine learning, we will be able to answer those questions better but for now, yeah, I'm not really sure.

[0:23:49.4] SC: What's kind of the future of your particular research both from the philosophical side as well as the machine learning side?

[0:23:55.2] MC: Yeah, we actually ended up talking about a sort of fairly small part of the research I've been doing which I think doesn't really –

[0:24:04.5] SC: Let's dive further into your research and –

[0:24:05.7] MC: Okay. One thing that was said at the symposium was that maybe at the moment, we're at a pre-Copernican revolution for our understanding of intelligence. Obviously in the revolution, we went from having humans at the center of the universe with everything revolving around it to humans as no longer at the center.

Our understanding of intelligence seems very human centric at the moment or at least the lay person or the everyday understanding. We're seeing a sort of move away from this whether the ideas of plant cognition and also this ideas of AI systems whereas Dennis was saying at the symposium. Alpha zero playing chess played very alien type of chess to him, it wasn't a human like way of playing.

It was a new type of playing. When we explored this intelligence landscape, humans are going to be a very tiny part of that giant landscape and with AI because everything is artificial, we have the ability to explore way beyond the scope of this little area that – well it is a very tiny area that biological life potentially exists in this landscape and as an even smaller area that human life exists in this state. But my particular interest and I think the big question moving forward for AI is when will we create intelligences and what type of intelligences has some kind of moral patience that you'd have some kind of –

[0:25:27.6] SC: Some kind of what?

[0:25:28.2] MC: Moral patience. So they are patience in our moral understanding of them so it can be ethically correct or wrong to put them in certain situations. So for example I think Bostrom have this term the mind crime where potentially we could create conscious artificial entities into some kind of slavery because they are just doing tasks for us or we could create entities that just live a life of suffering. They are never achieving their goals and it actually means something to say that they are suffering.

I think we need to explore this space of intelligence and compare it to the space of possible minds. The type of intelligences, not all intelligences are minds. We know that, it seems obvious but some of them are and we don't know right now whether it is just this little tiny and it seems absurd, it will be very pre-Copernican and to say, "It's just this little human dot that is the space in possible minds" where we need to map that onto the space of both what time it is. So I see that as the broad research goggles most important.

[0:26:30.8] SC: So beyond the tiny piece of it we discover what are some of the other kind of concrete research area is within the broad umbrella.

[0:26:39.9] MC: So I think that, well the most concrete question at least for me is how can we understand intelligence in a way where we can then say about certain agents that are intelligent whether or not they have a mind but obviously that itself is a massively huge question and it's going to take results from philosophy from neuroscience to get to the human understanding of like we know humans have minds.

So we can learn about minds from them and then from AI and from the completely other side of the field to think about what different types of alien intelligences is or artificial intelligence it could have minds.

[0:27:14.5] SC: Do we even have a functional definition of mind?

[0:27:17.3] MC: We don't even have a functional definition of intelligence that is one of the things. That was one of the great results I saw of this symposium was. We have invited all of these people together to talk about intelligence in different ways and they brought their own expertise but that expertise, each comes with it's own assumptions about what intelligence is and we even had this debate there with Alpha Zero now playing go and chess side of the humans.

If I told you I had a friend that he is a really good chess player, you don't actually think he's intelligent but now there is people saying, "Oh Alpha Zero is not intelligent. Chess is easy," which I mean there is not even an agreement on that front of what intelligence is. So yes, minds is going to be harder than intelligence. We haven't solved intelligence yet.

[0:28:01.8] SC: From the various folks involved in the symposium, are there patterns in the way they define intelligence that are easy to characterize?

[0:28:11.2] MC: I think so yes.

[0:28:12.0] SC: Is there like a page of Domingo's tribes of AI? Is there an intelligence version of that?

[0:28:17.6] MC: Yeah, I think there's definitely a camp that is very interested in human-like intelligence and what they tend to do is define intelligence in terms of human intelligence and then automatically assume that AGI as we move to more general intelligence is going to be on a path towards human-like intelligence because that is the type they are very interested in and then on the other side of the spectrum, you've got this idea that intelligence is much more broader than as humans as we could possibly imagine so I think you have these two separate camps.

[0:28:45.7] SC: Okay, interesting and so maybe circling back to the future and how you push all of these forward, how are you thinking about that today having just finished your pulling together the symposium and bringing together some of the folks that are pushing those research forward?

[0:29:05.3] MC: So I am thinking that it's never too early to start asking these questions. It may be too early to have concrete answers to these questions but it is definitely the time that we can actually bring these different types of people together to have this conversation because although everyone has different understanding of intelligence, we are getting results in these different fields that are comparable and we can start comparing them and talking about the issues.

So my feeling is, very optimistic that this is the important and right area that we should be working on and that we are going to get results now that we are at a state where AI, is as advanced as it is and our understanding of intelligence across the animal kingdom is that it is, but we can start bringing these things.

[0:29:49.7] SC: Awesome, well Matt, thanks so much for taking the time to chat with me about what you're up to. I wish I had the opportunity to attend the symposium. It sounds excellent and I know that you had some really excellent speakers and participants so I am looking forward to keeping up with the work of the group. Thank you.

[0:30:06.4] MC: Thank you.

[END OF INTERVIEW]

[0:30:12.1] SC: All right everyone that's our show for today. Thanks so much for listening and for your continued feedback and support. For more information on Mathew or any of the topics covered in this episode, head on over to twimlai.com/talk/91. To follow along with the NIPS series, visit twimlai.com/nips2017. To enter our TWIML one mil contest visit twimlai.com/twiml1mil. Of course, we'd be delighted to hear from you either via a comment on the show notes page or via a tweet to @twimlai or @samcharrington.

Thanks once again to Intel Nirvana for their sponsorship of this series. To learn more about the Intel Nirvana NNP and the other things Intel has been up to in the AI arena, visit intelnirvana.com. As I mentioned a few weeks back this will be our final series of shows for the year. So take your time and take it all in and get caught up on any of the old pods you have been saving up. Happy Holidays and Happy New Year. See you in 2018 and of course, thanks once again for listening and catch you next time.

[END]