EPISODE 102

[INTRODUCTION]

[0:00:10.8] SC: Hello and welcome to another episode of TWiML Talk, the podcast where I interview interesting people doing interesting things in machine learning and artificial intelligence. I'm your host, Sam Charrington.

Contest alert. This week we have a jam-packed intro, including a new contest we're launching, so please bear with me, you don't want to miss this one. First a bit about this week's shows; as you may know, I spent a few days at CES earlier this month. While there, I spoke with a bunch of folks applying AI in the consumer electronics industry and I'm including you in those conversations via these series of shows.

Stay tuned as we explore some of the very cool ways that machine learning and AI are being used to enhance our everyday lives. This includes work being done at Anki, who built Cozmo, the cutest little computer vision powered robot. Lighthouse, who's smart home security camera combines 3-D sensing with deep learning and NLP. Intel, who is using the single-shot multi-box image detection algorithm to personalize video feeds for the Ferrari Challenge North America.

Firstbeat, a company whose machine learning algorithms analyze your heartbeat data to provide personalized insights into stress, exercise and sleep patterns. Reality AI and [inaudible 0:01:34.0] who have partnered to bring machine learning based adaptive driving beams, or automatically adjusting high beams to the US. Last but not least, Aerial.ai who applies sophisticated analytics to Wi-Fi signals to enable some really interesting home automation and healthcare applications.

Now as if six amazing interviews wasn't enough, a few of these companies have been so kind as to provide us with products for you the TWiML community. In keeping with the theme of the series, our contest will be a little different this time. To enter, we want to hear from you about the role AI is playing in your home and personal life and where you see it going.

Just head on over to twimlai.com/myaicontest. Fire up your webcam or smartphone camera and tell us your story in two minutes or less. We'll post the videos to YouTube and the video with the most likes wins their choice of great prizes, including an Anki Cozmo, a Lighthouse smart home camera and more. Submissions will be taken until February 11th and voting will remain open until February 18th. Good luck.

Before we dive into today's show, I like to thank our friends at Intel AI for their continued support of this podcast. Intel was extremely active at this year's CES, with a bunch of AI, autonomous driving and VR-related announcements.

One of the more interesting partnerships they announced was a collaboration with the Ferrari Challenge North America race series. Along with the folks at Ferrari Challenge, Intel AI aspires to make the race viewing experience more personalized, by using deep computer vision to detect and monitor individual race cars via camera feeds, and allow views to choose the specific car's feeds that they like to watch.

Look for my conversation with Intel's Andy Keller and Emile Chin-Dickey later in the series for an in-depth discussion about this project. Be sure to visit ai.intel.com, where you'll find Andy's technical blog post on the topic.

Now a bit about today's show; in this episode, I'm joined by Andrew Stein, Computer Vision Engineer at Anki, and his partner in crime Cozmo, a toy robot with tons of personality. Andrew joined me during the hustle and bustle of CES a few weeks ago to give me some insight into how Cozmo works, plays and learns and he's different from other consumer robots you may know, like the Roomba.

We discussed the different types of algorithms that help power Cozmo, like facial identification, 3-D post recognition, reasoning, and even some simple emotional AI. We also cover Cozmo's functionality and programmability, including a cool feature called code lab. This was a really fun interview and you should also check out the companion video on YouTube starring Cozmo himself, which of course we'll link to in the show notes page.

Now, on to the show.

[INTERVIEW]

[0:04:42.5] SC: All right, everyone. I am here at CES and I am with Andrew Stein. Andrew is a Computer Vision Engineer at Anki. Anki is, well I'll let Andrew tell you all about Anki. Andrew, welcome to This Week in Machine Learning and AI.

[0:04:57.3] AS: Thanks. Thanks. It's cool to be here. Yeah, I can tell you a little bit about Anki's background and then a little bit about our products and Cozmo specifically. Anki is a consumer robotics company, and we currently have two products that are both in the entertainment space. One is Anki Overdrive, which is a car racing game. It's been out for several years. You can control the cars from your phone and they can drive themselves. You play against them like you would play in a video game, but instead of looking at a screen, you've got actual cars driving around on a real track and you're living with it.

Along those same lines, which in some sense is bringing a physical product to life, Cozmo is a robot character like you would see on the big screen, but brought to life in real. The goal was to really try to take this little robot character that the kinds of things we've seen in movies, but actually make a real one. I think that's a core tenet of the company is trying to bring physical products to life, trying to deliver on this promise of robotics and AI and consumer products.

Specifically, Cozmo is very focused on character and personality. You can play little games with him, he can recognize your face, he can play little games – if you leave him alone, he can do his own thing. He has three little cubes that he can carry around and make stacks out of and have lights on them. You can play games with them.

[0:06:17.4] SC: Also it gets a little feisty, doesn't it?

[0:06:19.3] AS: Yes. He has a lot of personality and that is a big part of it. I say that we're half core robotics company with all the tech that goes with robotics, which is very multi-disciplinary. It brings together a lot of different disciplines. We all add to that whole team of animators and character designers who are focused on the character of Cozmo, who is he and what does he – what is this personality and what does he like, and that's another big part of the company and the experience. I think it's what's cool about the company is bringing those two sides of things together.

[0:06:52.9] SC: Wow. Wow. You work on computer vision. Can you tell us a little bit about your background and how you got involved in CV?

[0:07:00.0] AS: Sure. Going way back in undergrad, for whatever reason I took a class, I think it might have been a graduate-level class, but it sounded cool on computer vision. I really liked the professor. I ended up working with him as a undergraduate researcher in state and did a master's degree there at Georgia Tech.

I had always enjoyed working with robotics. I actually had a job in high school doing robotics for sorting garments, hangered garments actually with giant industrial robots. Those two are things that I think both struck a chord with me. Then went to Carnegie Mellon to pursue my PhD in robotics focused on computer vision there.

[0:07:37.6] SC: Okay. How does computer vision fit into Cozmo? Cozmo is so small, I don't even see a camera just anywhere.

[0:07:45.8] AS: His camera is actually in his face. If you look at him, the little hole there that looks like a mouth is his camera. Creepily enough, he has his eye in his mouth. Yeah, it's a pretty core piece of the robot, because it's really his main input, his main source of sensor input. Cameras are by their nature are very data-rich source of information. They're also very inexpensive, so that's a good combination.

That's his primary way of sensing the world is through vision. That's how he knows where his cubes are, he can see them. He has to make their poses in three dimensions very accurately, so that he can pick them up and stack them. He can perceive motion and he also sees faces, both human faces and cat and dog faces. Beyond detecting faces, he can also learn to recognize human faces, so you can teach him your name and he will remember you.

[0:08:36.1] SC: Wow. What's some of the technologies that go into making this happen from a CV and an algorithmic perspective?

[0:08:43.9] AS: Sure. I mean, face detection is certainly a big one; face detection and face recognition. There is a lot of proprietary stuff around how we do the 3-D pose estimation of the cubes. There is intelligent reasoning about – given the geometry we know of the robot and the specifics of the camera, which we know the intrinsic parameters of the camera, we can do things like reason about the ground plane in front of the robot, even though he doesn't have a depth sensor, we can start reasoning about the ground plane in front of him, which turns out to be pretty powerful.

Then we'll layer on top of this, all these – lots of other technologies and coding, sort of low-level motor controls, path planning. Haven't used the word AI, or the acronym AI; it's so overloaded at this point, but sort of the AI behind how Cozmo models his emotional state and how that drives, what behavior he chooses to do at any moment. Like I said, he is doing on his own. You're not remote controlling this robot. He's some character. What makes him decide to do A or B at any given moment and how do we keep that making sense and can't just be random and also can't be scripted. How to drive that behavior system.

[0:09:51.8] SC: Well, let's maybe start with the computer vision stuff. There are a lots of ways to do computer vision traditional stuff and convolutional neural nets. They're obviously very popular. When I look at this thing and think about the prize point power, stuff like that, I'm guessing that you're not running neural nets in there?

[0:10:11.6] AS: Well, there's another reason for that actually, which people tend to forget given how popular they are in the news now. This product actually started before we even launched Drive, which was back in 2013. I was the first person working on the product. There was nobody else doing anything yet. There was no code yet. There was also no really – okay, so I shouldn't say there was no deep learning, because neural nets have been there on time. The revolution, if you will, hadn't occurred yet.

[0:10:35.9] SC: We forget that we're still at the –

[0:10:37.9] AS: It's so early.

[0:10:38.4] SC: That's right.

[0:10:39.9] AS: TensorFlow didn't exist yet. All these things people are so familiar with now. It wasn't even around yet. Yeah, I'm what's known as a classic literate in the computer vision engineering, which is ridiculous.

Yeah, when we started all these and started picking hardware and nailing down price points and what processing power he was going to have, we actually did start to do a lot of the vision on board, but again with more classical techniques. For face detection, things more like Viola-Jones face detection.

[0:11:08.9] SC: Like what?

[0:11:09.5] AS: Viola-Jones. It's a classical way of doing face –

[0:11:11.3] SC: Viola-Jones. What is that?

[0:11:13.2] AS: It's a means of progressively filtering an image with more and more very, very simple – simply design filters that are very, very fast in order to – via a cascade sort of rule things out slowly over time, but be very efficient and eventually learn the pattern in the image. It's basically looking at local contrast, patterns in the image that look like a face. Works very well and I'd still often use today. That's the thing we would use.

[0:11:39.8] SC: The idea with that face or no face, or is it what allows you to identify individual people? Can you –

[0:11:45.2] AS: Yeah. That is face or no face. That's what I would call face detection, then I would contrast that with face recognition, which is given a face who is it exactly? Anyway, we started by doing that in the marker detection trying to do it on the robot. We were able to get actually quite far with that, but at some point realized, "Okay, this is just too limiting." We always knew there would be a companion out the same way that there is Overdrive.

At some point we decided, "All right, we're going to take the plunge and just put all the smarts really in the device." The way the product works is that you have an app that connects to – you connect your device to Cozmo's Wi-Fi Hotspot, and he's actually streaming his images to your device and all the computer vision path planning etc., is actually happening on your device.

[0:12:29.2] SC: Interesting.

[0:12:30.4] AS: That's what again, like as you pointed out, that's what allows us to sell it at a price point and a scale that we're able to. Otherwise, especially given hardware from three plus years ago, there's just no way we could've gotten all the capability on the robot.

[0:12:44.7] SC: Right, right. Is there some limited ability to operate if the – say your iPad runs out of battery or something like, is it able to go into some autonomous mode?

[0:12:54.9] AS: Very, very little. He will shut himself down. We try not to just have him die, but it really is – it is quite tied to the device, because so much what was there. All the animation in fact, so we talked about the animation, but the animations that play on him, the sound, his facial animations are also actually all stored on the device. While those are streaming to him, his images are streaming bac to the device, so there's a lot of data going back and forth. I should say it's all within that network. There's no cloud anything. This is a closed network.

[0:13:24.7] SC: That's Bluetooth, or –

[0:13:26.0] AS: That's pure Wi-Fi. It's pure Wi-Fi.

[0:13:27.0] SC: Pure Wi-Fi.

[0:13:27.7] AS: Yeah, the problem with Bluetooth, which is actually what we use for Overdrive is just bandwidth. We couldn't stream the images at full frame rate over Bluetooth. The robot controls we could, but the actual image data we couldn't.

[0:13:39.9] SC: Okay. You started looking at for the image detection the Viola-Norbert –

[0:13:45.9] AS: Viola-Jones.

[0:13:46.7] SC: Viola-Jones.

[0:13:48.0] AS: I like Norbert. I don't know.

[0:13:48.2] SC: I'm thinking of Norbert. Oh, man. Yeah.

[0:13:52.8] AS: That's a new algorithm. We have to develop it ourselves.

[0:13:56.6] SC: Is that what you ended up doing on the device? Or now that you have access to the device, you're able to do more sophisticated things?

[0:14:05.9] AS: I mean, without going in too much detail, that's basically what's running for the detections. Cubes is a completely different thing. Those are detected via certain proprietary method that both allows us to detect the cubes and then estimate their pose in 3-D. Again, that's super important, his little fingers that he has to get into little slots to pick up the cubes, yeah there's only a couple millimeters of spot there.

We've got a robot driving around on treads and treads are really hard to model. The way he moves, we really have to be able to get feedback constantly about where the cube actually is so that we can drive him accurately and pick up the cubes. That was just a huge part of the project for a very long time was just how do we make this thing pick up cubes?

[0:14:46.6] SC: Wow. Understanding that it's a proprietary approach, can you give us some analogies that help us understand what are the technical challenges beyond, obviously the precision that you just mentioned? What are some of the algorithmic approaches you looked at before you went down the path of needing to roll your own? What might you consider if you were starting again, that kind of thing?

[0:15:13.4] AS: The obvious thing if you look at these cubes is probably maybe QR code. It's along those lines. One big reason we didn't go QR code is the appearance. We are a tech company. We're building technical products, but there is a big design component to this, how the robot looks, how he behaves and what we wanted the cubes to look like. Stylistically nobody liked the QR codes. They scream with the wrong thing for the product.

One of the things we want to develop was a similar idea that allowed us to encode information on the sides of the cubes that gave them information, but that gave us aesthetic control over what they looked like. It's a similar idea to QR codes in some sense, but with a aesthetic component.

Then as far as estimating the 3-D pose, what it effectively comes down to is that we know points on the cube, points on the marker that are inside the cube, and we know the intrinsic calibration of the camera. It's basically about **[0:16:10.6]**. That's a whole other interesting issue. We have to calibrate every robot individually in the factory to get that.

Given those two things, we can see where those known 3-D locations on the cube project into the image. Once we find them in the image and we now – it's the 3-D shape they belong to, via that correspondence into math, you can back out where that 3-D object must be in order to project that pattern.

[0:16:36.8] SC: The 3-D points that you're referring to are –

[0:16:40.1] AS: Anything on the –

[0:16:41.2] SC: From decals, or –

[0:16:42.1] AS: Yeah. Anything we know. You could use anything really. But you just want a very accurate known position on the cube.

[0:16:48.6] SC: You've got these known graphics, we can call on codes and not QR codes. When you look at them, you think that these are just like graphical flourishes, but you look more closely and you can see that each of the sides is unique and they're conveying some information.

[0:17:05.3] AS: Yeah. People don't often notice that. It was one of the challenges of designing them is that we had competing goals. One was that we wanted all sides of the cube to look the same, so that this cube had one marking on it. But we also wanted Cozmo to be able to tell the difference from the different sides, because he can control the lights and we want to know which light he's turning for example.

[0:17:22.6] SC: Yeah, I didn't even notice that. This is your –

[0:17:24.4] AS: Yeah, there's four lights on top.

[0:17:26.9] SC: Okay. I was also noticing that this cube in the middle is like your paperclip cube. It's like the top of the cube.

[0:17:32.2] AS: That's exactly what I say, it was like it's a paperclip.

[0:17:33.7] SC: That one looks like a stack of things.

[0:17:36.0] AS: Everybody’s got their own what they see in these things. Yeah.

[0:17:38.6] SC: War shock you get or whatever.

[0:17:40.6] AS: Exactly. Exactly.

[0:17:41.9] SC: This one looks like a baby in a fetal pose.

[0:17:45.6] AS: Yeah, that’s a common one too. It is funny when people see these things. It was hard, because part of the design here was – I guess, if we wanted the aesthetic component, but we also at the time we were locking down making all the hardware, we didn’t necessarily commit ourselves to a particular meaning. If you made it the treasure chest cube, it’s like – just everything you do – does with the cube have to relate in some kind of a treasure chest.

We didn’t want to get too much iconography. Yeah, these are these general purpose designs. The challenge that we realized later is that exactly what you just experienced is how to refer to them is very typical. Customer care when they get a call about the cube or something they were always like – there is actually a number engraved in it that you can find, so you can actually figure out what it is.

[0:18:31.6] SC: Well, an obvious would’ve been like make them different colors. Did that mess up your algorithms or something, or?

[0:18:36.6] AS: No, no, a color would’ve been okay. I think it’s just from the design perspective. I think they wanted them to match more.

[0:18:43.3] SC: Interesting. I understand for the people that are listening that this is a very visual conversation.

[0:18:48.2] AS: I’ll be talking about computer vision.

[0:18:49.8] SC: Yeah, exactly. I’m going to at the very minimum include some pictures of what we’re talking about on the show notes page, but more likely or as well go into maybe shoot some video, this thing in action and post it up on our YouTube channel.

[0:19:08.1] AS: One thing that often I think surprises people after hearing it talked about, or referred to, I don’t think they realize how small he is.

[0:19:14.3] SC: I didn’t either. I was really surprised.

[0:19:15.4] AS: That’s why I like to bring one. Generally people, I think imagine them much larger. He’s actually fits in the size of the palm of your hand. He’s quite tiny. That’s for a couple of reasons. One is actually maybe not obvious, which is for him to have the personality he does and to move around as fast as he does to exhibit that personality and be cute and playful, he has to move quickly. If you build a heavy big robot, his lifter here moves too fast and you got your finger in there, you could actually hurt yourself.

In some sense, there’s a safety component to it. It’s also just part of his personality. It makes sense for him to be cute. It wouldn’t if his too big. I don’t remember, it was 40 or 60 or something di following design iterations on this thing. Some of them were much bigger and the smaller ones always win out. It just feel s right for him to be in the palm of your hand.

That in turn has impact on the way the sound design is done; what should he sound like given how big he is. The difficult side of course is really for the manufacturing and mechanical engineers to squeeze in all the 300 something parts into this tiny little robot. There is no free space inside that thing. It is packed.

[0:20:27.5] SC: I can imagine. I can imagine. We were talking about the size of this thing, of the cubes. You’ve got this – this algorithm is interesting. Each of the cube has a six size that has the QR code life thing that is a consistent design element for each cube. Then this thing can look at a cube and is probably relatively easy to identify if a cube is in the frame and then it can pick out the – infer the angle of the different size that it’s able to see. From that, figure out, you can create some model like a projection I’m thinking of that –

[0:21:13.5] AS: Yeah, that’s projected geometry is what it all comes down to. The projection of the 3-D points onto to the 2-D plane of the image. Given we know what the 3-D points are.

[0:21:21.5] SC: You can bag into a rotational, a center and a rotation.

[0:21:25.8] AS: Yup. That’s exactly right.

[0:21:26.8] SC: Then you would feed that into whatever, like classical control algorithms to make it move to the thing, lift it up, pick it?

[0:21:34.2] AS: Yeah. Plan a path into a known location with respect to that cube. Then yeah, it's a control problem of – as I'm driving forward, make sure I keep it in the right place. My fingers get in there and pick it up.

[0:21:45.6] SC: That seems amazingly sophisticated for this little cute thing.

[0:21:48.0] AS: Yeah, people don't understand how hard that is. One of the things that I think actually is pretty interesting about this is it's I think more for robotics geeks than the average consumer is that one of the things that holds robotics back from doing more and more I think consumer products is manipulation, is the ability to change the world around you.

It's a very, very hard problem both mechanically and from an AI standpoint, from a software standpoint. I would argue, this is the first little mobile manipulator, especially at this scale and this price point. I mean, Roombas are effectively a manipulator and that they suck up stuff. They are doing some work. But Cozmo can actually yeah, do this very hard problem of driving up, picking up a cube and then stacking on top of another cube. It's not an easy thing to do. It is not an easy thing to do.

The other half though that may not be obvious – it's definitely not obvious about the cubes, so not only that they have lights, they also have an accelerometer inside of them. Cozmo talks to them over a radio connection, which is like Bluetooth. That allows them to know via the accelerometer if the cube has moved. If I pick up the cube, he's aware.

What that means is that if – if I've seen the cube and I've estimated its 3-D pose with respect to the robot and the cube doesn't move, now I can also do the reverse, if I drive around in Cozmo, or I pick up Cozmo and –

[0:23:08.5] SC: You can locate Cozmo based on the cube –

[0:23:10.6] AS: Once he sees the cube again, if it doesn't move he now knows his position with respect to the cube, which means he now knows his position with respect to the old map he was building. It's something that I think people tend to forget is it's not just like pure stimulus response, “Oh, cube. Oh, face. Oh, whatever.” He’s remembering all these. He’s keeping up with the 3-D poses and the cubes. He’s keeping up of where you are in space once he sees you and there’s big reasons for that. It makes him look smarter. It allows him to do behaviors, which turn out to be super important.

For example, right before he decides he’s going to go pick up this cube and he knows you’re off to his right, he might stop and do the same thing a little kid does, which is look up and make sure you picture him. That little moment of eye contact makes it more about this interactive experience where you’re drawn in and you’re very aware that, “Oh, he knows I’m here.”

As opposed to, I’m just an expectator watching a robot pick up a cube. It’s like, “Oh, Cozmo knows I’m watching him.” That little bit of the mixture of that technical component, that technical capability with the character and personality I think is really good example of how this starts to fit together.

[0:24:19.5] SC: Interesting. You didn’t literally use air quotes when you said AI with relation to the personality, but clearly there is a connection between the way we think about AI and the idea of personality, like what are some of the – how are you doing that? What are some of the approaches to –

[0:24:45.1] AS: Personality component?

[0:24:46.4] SC: Yeah.

[0:24:47.4] AS: Yeah. Maybe I mentioned there is this notion of Cozmo having emotion. We do internally model his emotional state; how happy versus sad, how he is, how converse as anxious he is, how socialized versus lonely he is. I remember those all are the same words we actually use in the code. He effectively has this set of traits or properties, which are changing all the time and different things that happen to Cozmo affect them.

Another good example involving faces is that he’s designed to be a social friendly robot. He defines his personality to be one that enjoys being surrounded by people. If he’s driving around for a long time he doesn’t see anyone, at some point his loneliness may creep up. Once it gets high enough, it may trigger him to switch into a behavior, which is look for faces, because he’s lonely. That will change him to a mode where now he keeps his – keeping his head tilted up and he's looking around and he's not distracted by his cubes or whatever. He wants to find a person.

Sees a person and now that triggers an emotional change, where his socialization goes up and his loneliness goes back down. That allows him to switch out of the behavior and do something else. That's the idea of what I was referring to earlier of preventing it from being either just random behaviors, which over time you can tell as random, or being fully scripted, which also doesn't feel natural. It is in response to what's been happening and what is currently happening to him.

[0:26:20.3] SC: Is there a notion of like a long-term personality? Meaning, the thing that came to mind is like the Microsoft Tay chatbot. They got trained by Twitter to be a Nazi. If you ignore your Cozmo long enough, will it become permanently sad or something like that?

[0:26:42.8] AS: It's a very good example and you hit exactly the – The answer is no. You had exactly the reason. We were concerned, like what happens if you just – your robot ends up irreparably depressed? It's not really what we wanted. We felt like what we were trying to create was – there is a definition we have of who Cozmo is and what his personality is. We have character designers and that is their job. They are the owner of who is Cozmo, what are his motivations, these sorts of questions.

[0:27:10.1] SC: It's Cozmo. It's not my Cozmo. When I take it to get out of the box, there's not like a random seed that, or something on maybe more continuous than a random seed, but that says, "This is my Cozmo's personality." It's more Cozmo's personality.

[0:27:24.4] AS: I think that's something we're just exploring at this point. What I would call the personality is more fixed and we're in more control of that. The mood I would call, which is more transient is what you're controlling by what you do in them. If it keeps falling on the floor, or yeah if he doesn't see anybody for a while, again he might get lonely, those sorts of things. But they don't exhibit a long-term effect, because it can be very hard to control like, "Wait. Where does that go?" Yeah, we I think we're a little cautious about that.

[0:27:55.6] SC: Okay. On one end you've got random on another and you've got totally scripted. I'm imagining somewhere in the middle is like a state machine that's sufficiently complex, that it doesn't seem like either of the two. Is it like that approach?

[0:28:10.2] AS: Yeah, I think that's a fair comparison. There are predefined behaviors and for example, games that he can play, which themselves are a little state machines, which are very much engineered. They look for faces. We didn't learn that behavior of how to look for – Okay, what it means is he look up and look around in this way and those things are tuned. Yeah, I think looking at as a state machine, where the transitions between states are very much driven by not only his sensor input, but also this underlying emotion engine is when you think about it.

[0:28:46.0] SC: I guess, one of the questions that jumps out to me as a geek I guess is like, is this thing programmable? Can I like –

[0:28:52.5] AS: Awesome question.

[0:28:53.9] SC: Can I use it as an experimentation platform?

[0:28:56.7] AS: Absolutely. That is actually one thing I'm super excited about that we have done with the product and actually we did from day one. When we launched it, Cozmo comes with an STK. In fact, not everyone realizes it, but in the app that you use to talk to Cozmo, if you go to settings, you scroll over there is a button which says enable STK. Everybody has this out of the box.

You enable the STK, you plug your device into your computer or US and then you can program them by phone. It is an extremely full feature and ever and expanding STK. It's actually totally incredible all the things you can do. Having done a PhD on robotics on research platforms which cost tens of thousands of dollars are usually broken.

The fact that this $179 robot allows you to do totally well of emotion control, or motor control, all the way up to just people on face recognition and just use it and path finding is crazy. It's designed for six-year-olds, so it can fall off the table and not break.

It is an awesome programming platform. It's actually being used both at Carnegie Mellon and Georgia Tech, I guess both my alma mater. In their programming classes, both I've been grad in graduate level. We've got some cool – I can't think of the name right now, of programming camps in the summer. They're starting your Cozmo as the platform.

In concert with all these, so I've mentioned the STK. That's full-blown geek level robotics program you can do as – well not an expert, but someone at a graduate, or undergraduate level or someone who really knows Python. People by the way are also writing – creating movies with him online by scripting with the STK, which is really cute.

[0:30:38.4] SC: Really? That sounds like fan-fiction genre.

[0:30:40.5] AS: There's actually a whole YouTube channels. Yeah, life with Cozmo is one worth checking out. It's really, really creative videos done with him. Super impressive stuff and it's all through the STK.

It's really cool to see both used for research and also for creative outlets like that. Beyond that, there's a whole bunch of other stuff, which scratch-based programming. Scratch is a dragon drop block-based visual programming language developed by MIT and Google. Last summer, we actually released a early version of that where you could effectively sequence the robots. You got to view very basic blocks and you could do things like drive straight, turn right, look up. You could do fun things like wait until you see a face smiling.

You could actually have Cozmo do drive straight, look up and then sit there and then once he saw a face smiling he would proceed to the next block. You can do fun little programs like that. That was meant for the other end of the spectrum, people who have never written a line of code, have no idea about it. It teaches you how to break a problem down into steps, how to write sequences, how to sequence that and some of the basics that you can do with a robot.

At that point we had the very beginning end of the spectrum and the gadget-level, programming-level end of the spectrum. Last fall, we actually released what we call code lab, which took that early version of scratch and added to it another mode, which is more advanced. We have what we call now sandbox mode and constructor mode. For listeners who those, those basically correspond to horizontal and vertical in scratch. Horizontal is sequencing and vertical as it actually be branching and loops in more complex structures, but still visually.

We took code lab constructor and really basically enabled almost anything you could do in the STK, but in dragon drop programming with both. The first time I used it I was blown away how easy it was and how much you could build, how quickly you can do math in there, or you can do lot of operations. You can really do what it's got; trig functions. You can do whatever you want.

Now we have this full spectrum of very beginner level, drag and drop sequencing to really full-blown programming but still with dragon drop blocks, so that you can see how that works. Then once you're comfortable there, you could easily transition into Python and understand how to write code there.

It's a really nice transition. To move that along, in code lab we've also released these feature projects and we are continuing to do that. It's cool, because we can build a little fun activity with code lab. In the app it comes up as a little icon. You can open it up, oh this sounds fun. You can play it. It's a little game, or a little activity like making Cozmo play different instruments when you tap on the blocks for example, all written in the code lab.

The cool thing is there is a button on all of it as the inside. You click that button and then it actually shows you the full scratch-block based program and you can see like, "Oh, now I see how they did that." You can customize it or whatever, but it's like the old way we all learned the right webpage is to do the source. In fact, how I see how they did that. It's again a really cool way to dive in and get some ideas for what's possible.

[0:33:53.7] SC: Now with the STK and the scratch piece, you mentioned a motor control. You can also get a feed of the images and like try to, so you want to play with your own facial recognition algorithm, or –

[0:34:10.1] AS: Yeah. Totally can manage for you. The STK, yeah you can get the image feed. It's a little harder to do that in scratch to display it. That's something that I think is worth exploring. Yes, the STK absolutely. We've had people do that. People who are computer vision researchers in grad school who want a robot, they don't want to deal with the path pointing part of it. I found a thing and that with an obstacle in 3-D and I want to drive a path around it. I'm focused on the vision. I don't care about the path finding. They could use the path finding, but they get their image feed and they can do their own detection.

There's been interesting work, like taking Cozmo's image feed and running it through some of the popular deep learning networks and learning to recognize objects and things. Again, they have the power of a whole laptop to run it on. Yeah, we've seen people do some really interesting projects. We have a very active developer form on our website. People post this kind of stuff all the time and we've had really, really great response.

[0:34:59.9] SC: Interesting. What's next for either this product or the company? Is it building on this as a platform, or coming out with the next robot, or the next thing?

[0:35:12.9] AS: Yes, yes and yes. There's not so much I can say about too far down the road, but I will say yes, we're adding – we're definitely expanding Cozmo's capabilities to be able to see and understand more. A lot of the user-facing stuff in the near future is focused around code lab.

One of the benefits of having this code lab universe where we have these projects is that is also makes it easier for us to release new content. We don't need a C++ developer who knows all about robotics to write a new little fun activity.

We can have a designer, a game developer who may not be as much of a hardcore coder, but has a cool idea and they can drag in Dropbox and make a project and that can be part of the app. We can also take user content, so you can send us your content, you can share your projects and cool stuff that actually you need. We might actually deploy with the next version of the app.

There's a lot of stuff around code lab coming. There are new things coming in the Cozmo product, I can't say too much about. Then I guess long-term, I would go back to saying how at the beginning, I said we're a consumer robotics company. I didn't say we were a toy company. We're currently focused in entertainment and that's very deliberate for a couple of reasons. One, we felt that at the – to develop the capabilities you needed both technically and from a manufacturing scale and price point perspective, this was a good place for us to start building the actual product that we could sell and market and build a successful company on, as opposed to jumping to the far end of like, we're going to have a humanoid in your home and it's going to clean your house.

Yeah, and how are we going to fund that company, right? I'm trying to keep an eye on building a business at the same time and how to take steps, sort of build products and stepping stones as we build out core technologies and core capabilities to get to those big fancy robots everybody wants.

[0:37:08.6] SC: It seems like a lot of the companies in this space take that approach in some way shape or form, like iRobots we know them for the vacuum cleaner, but they've got a lot of government robots and defense robots. I'm sure they're eyeing , the home robotics market in this like grows and creates opportunities.

[0:37:28.9] AS: Yeah. I think that that – I think it's a necessary thing people often – that everyone sees the stuff in movies and TV right that's what I want. Despite all the headlines about AI and deep learning etc., it's still a long way off. I think being careful about building that technology out in a very deliberate manner and creating products along the way that make good products themselves support it.

Because to us, a robot is not a product. It is a technology. It is encapsulation of technologies, together which make a product, but you still need a product idea. I think to that end, it's not only are we trying to build those technologies. The other thing that we feel is important, we like to say not only is the IQ of the robot important, the techy, smarts AI, but the EQ is also important. We're going to build these robots and they're going to be living in our homes and that is our goal. We want a robot in every home.

Those robots, we don't want them to be weird appliances that sit off in the corner that it's this thing you don't interact with. We've seen this with Cozmo. It's a really interesting moment when you make eye contact with the robot, or you would assign a personality to it or have a bond with it and we definitely see it with this little robot. It's a whole different experience, and so I think that that expertise we're building in how to take things we know about movies and character design and deploy them in hardware and deal with that side of the human robot interaction piece of the puzzle is also super important, and I think will be important for all our products in the future.

[0:39:06.4] SC: Are there any learnings that you can encapsulate for us on I guess the intersection of AI and consumer electronics, or like the challenges of putting AI and consumer electronics? In this podcast, we talk a lot about enterprising stuff and industrial robots and things like that. I'm wondering about the specifics of AI and games and entertainment, toys and electronics and that kind of thing.

[0:39:37.9] AS: Yeah. I think a few things; consumers don't actually care about – I think your listeners do and I do, but at large, consumers don't care about the actual tech. They just want it to work and be cool. We all have our own cellphones in our pocket. How many people have any idea how that technology works, right? It's ridiculous what we do every day on our cellphones, but people just want it to do all that awesome stuff.

I think that is one thing is as engineers working on their product, I have to remember computer vision is not the product. There is a product and it has goals and computer vision is in service of those goals, not the other way around. We often don't use the latest and greatest tech or idea, because it's like well, can these users going to be able to tell if we're doing that? What is the actual end result of using that technology?

I think keeping in mind that in the consumer space you're building a consumer product. You're not necessarily building a technology that's B to B and will be using other products. Keeping that end goal in mind is probably one of the big ones.

I think another big thing about using AI in particularly everything over the last N years is about probabilistic reasoning effectively. People want a yes or a no, or a guarantee that it's going to work. If these cases didn't work, work in those cases. If there's anything we know it's that it's very hard to nail that down. You can say it's based on our data it's going to work 95% of the time. Well, like enumerating the 5% of cases, you can't do it.

A lot of what we spend our time doing is it's easy to get the early prototype of the cool behavior. It's what do you do in the weird edge case and 5% of the time that it doesn't work situations? All those educations are really complicating. We have kid picks up a robot right in the middle of the behavior X or animation Y. It's like, "Okay, wait. What happens then?" Just enumerating all possible states in the state machine is not really a viable solution either.

I think edge case handling is a big, big deal when you start trying to deploy these things that you know will have failures or have false positives. How do you incorporate that into the product as opposed to pretending it doesn't exist? Because it will happen.

[0:41:51.9] SC: Have you developed any methodology for tackling that specific issue, or is it each behavior, each edge case is different and it's just knowing that you need to think that through that is –

[0:42:05.9] AS: That's a good question. Particularly, there are guys that work on more specifically and focused on that behavior system. I would say a little bit of both. Over time, the way that our behaviors encode or engineered are designed to handle things better naturally just by virtue of the system architecture is set up.

There are ways to build what we have learned I think into the system, but there is a lot of that secret sauce, black magic thing with deep learning. There is things I feel like people can't quite explain yet, this sort of like I've just done this and I know what is and isn't going to happen. Some of that I think is just at this point, our internal knowledge of how it works. But yeah, I think actually over time as you start to codify what those things are, there are definitely places in the code where the architecture again supports that or makes it easier or it handles things for you that you realized this always happens, or you need to wait and just automatically detect and handle that.

[0:43:06.6] SC: Okay. Anything else on your things think about from a consumer products perspective?

[0:43:14.4] AS: I think those are probably the big ones. I guess the other one, given that data is such a huge thing, for training all these models and labeling is such a huge thing. I think for robots in particular, images in particular within that, getting training data is I think even harder on robots, because the degree to which the robots view the world and images mined from the web differ is huge.

The statistics of the data that a robot sees, it's all motion blurry or terribly exposed, or like half your arm or whatever. Nobody has actually pointed the camera at something and taken a picture. I think people tend to forget that there is already some selection bias in mining images from the web, or images from Facebook, or somebody how it became and took the photo and framed the shot.

[0:44:05.0] SC: Decided to upload.

[0:44:06.2] AS: Decided to upload. Exactly. Very true. Cozmo is driving around, taking images all the time. You just get weird random garbage all the time and terrible exposures and bad light balance and lots of motion blur and a weird perspective. He's looking up at the world. Nobody takes pictures from there.

Gathering that data is a big challenge and I think it's not to be underestimated how much it matters to try to get data appropriate for your problem when it's robotics and not something else.

[0:44:39.0] SC: It seems like that would be easy. You just make a hundred of these and throw a bunch of blocks around and have them run around and shoot a bunch of video.

[0:44:46.7] AS: I mean, it depends on the problem you're trying to solve. If it's block, you can probably design a scenario. You're right. Some situations –

[0:44:53.5] SC: If it's the people interaction thing that's –

[0:44:55.8] AS: Yeah, people vary. Varies a whole lot. All these things are challenging, just different types of rooms. We're building an office and the office environment looks very different from people's homes. We also don't – for privacy reasons, we're not just going to gather data from people's homes and upload it to our server. Yeah, the data collection problem is a big one.

[0:45:16.7] SC: Yeah, makes sense. Well Andrew, this has been a great conversation. What I'm going to do now is I'm going to hit pause and go grab my camera and we'll let you fire this thing up and see it in action.

For the folks that are listening on the podcast, they may not catch this part, but jump over to our YouTube channel and you'll check this out. For those who aren't going to do that or will be doing that later once we're off the train or whatever, thank you so much for taking the time to chat with me.

[0:45:49.6] AS: Sure. No, it's been fun. Lots of good questions.

[0:45:51.6] SC: Awesome.

[END OF INTERVIEW]

[0:45:56.3] SC: All right everyone, that's our show for today. Thanks so much for listening and for your continued feedback and support. Remember, for your chance to win in our AI at home giveaway head on over to twimlai.com/myaiconteste for complete details.

For more information on Andrew, Cozmo or any of the topics covered in this episode, head on over to twimlai.com/talk/102. Thanks once again to Intel AI for their sponsorship of this series. To learn more about their partnership with Ferrari North America Challenge and the other things they've been up to, visit ai.intel.com.

Of course, we'd be delighted to hear from you either via a comment on the show notes page, or via Twitter directly to me at @samcharrington, or to the show at @twimlai.

Thanks once again for listening and catch you next time.

[END]