EPISODE 52

[INTRODUCTION]

[0:00:10.5] SC: 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.

The show you're about to hear is part of a series of shows recorded in San Francisco at the Artificial Intelligence Conference which was hosted by our friends at O'Riley and Intel Nervana. In addition to their support for the vent itself, Intel Nervana is also our sponsor for this series of podcasts from the event.

A huge thanks to them for their continued support of this show. Make sure you check out my interview with Naveen Rao, VP and GM of Intel's AI products group and Scott Apeland, director of Intel's developer network which you can find at twimlai.com/talk/51. At the AI conference, Intel Nervana announced a dev cloud, a cloud hosted hardware and software platform for learning, sandboxing and accelerating the development of AI solutions.

The dev cloud will be available to 200,000 developers, researchers, academics and startups via the Intel Nervana AI Academy this month. For more information on the dev cloud or the AI Academy, visit intelnervana.com/devcloud.

My guest for this show is Paul Tepper, Worldwide Head of Cognitive Innovation and Product Manager for machine learning and AI at Nuance Communications. Paul gave a talk at the conference on critical factors and building successful AI powered conversational interfaces. We cover this and a bunch of other topics like voice UI design, behavioral biometrics and other interesting things that Nuance has in the works.

Now, on to the show.

[INTERVIEW]

[0:02:10.5] SC: All right everyone, I am here at the AI conference in San Francisco and I'm with Paul Tepper who is the Worldwide Head of Cognitive Innovation and Product Manager for machine learning and AI at Nuance Communications, the enterprise division of Nuance Communications in particular. I had the pleasure of meeting Paul at the AI conference in New York just a few months ago and he was kind enough to volunteer to jump in the hot seat. Welcome Paul.

[0:02:39.6] PT: Thank you, good to be here.

[0:02:41.4] SC: Absolutely. Why don't we start by having you introduce yourself to the audience and talk a little bit about your background and how you got into machine learning and AI?

[0:02:52.8] PT: Sure, well for about a year now, I've worked at Nuance Communications, in our enterprise division. I lead a team that has a few functions, one of our main functions is identifying high value problems for which the company doesn't yet have a solution and working with our large corporate research division to see if we have new forward looking research that could be sort of productized or prototype to get it for the customers.

As a way to kind of move innovation and product. Give the product manager and teams across the company new opportunities and things to look at there. I did my PHD at North Western, my focus is on competition linguistics, particularly in dialogue, I did a lot of work on none verbal behavior, gesture in particular and I also spent a few years prior to Nuance working at a startup, it's not around anymore, it was called Eddy Bon and we focused on building a cloud based NLP platform that really focused on human and loop computing and crowd sourcing as a way to quickly build data sets out, to build custom models for NLP.

[0:03:53.2] SC: Okay, I don't think I realized the North Western connection. I'm a PHD dropout.

[0:03:58.7] PT: Congrats.

[0:04:00.4] SC: Yeah, engineering on my side. Were you engineering or?

[0:04:03.2] PT: Yeah, I helped found this program called The Technology and Social Behavior Program that was a joint degree in computer science and communication studies of all things but that's where we ended up in communication studies doing a lot of work at that time.

[0:04:16.4] SC: Okay. Super interesting. Do you miss Eviston?

[0:04:20.4] PT: I'm more of an East Coast guy. I undergrad at Rockers and I spent a year in Scotland, did my masters out there so I've traveled around quite a bit but I'm definitely an East Coast guy.

[0:04:30.8] SC: Nice. I think the folks will get a little bit more about what you're up to now if you maybe spend a few minutes talking about Nuance and what Nuance is doing.

[0:04:41.6] PT: Yeah, Nuance is a fairly large, complicated company, have 14,000 employees and several different divisions, they were division focuses strictly on mobility, mobile products, automotive products so we have ASR and a lot of cars that people drive and do recognition for cars, the end car systems.

We've got a healthcare division that is top systems for dictation for doctors to do electronic medical records, imaging division. In my division particular focus on enterprise communications and we have two main product areas or areas of focus and that's one IVR systems or interactive voice response and that was one of the first commercial applications of speech recognition.

Systems that you can call in to and instead of doing kind of dial tone menus, you can just say what you want, conversationally and more recently, our focus has been on the doodle side. Now they're called chat bots but we've been calling them virtual systems for a long time.

Actually, there's a lot of overlap in those two technologies. A lot of the technology that has been built out for IVR. So, to recognize intent, to recognize concepts and do antique instruction inside sentences and then also a dialogue build out the dialogue for the graphs or trees to build through a dialogue.

A lot of that actually works similarly in the chat world. However, in the IVR world, the audiences tend to be a lot shorter, the tower bold has talked a lot longer so different complexities there, that's actually an active area of research for us. How do we use our user experience, how do we use UX to get people to talk longer and IVR still give more.

The natural language understanding or NLU as we call it has really far out paced what people actually say today in IVR systems. That's an interesting technical challenge for us, yeah.

[0:06:21.9] SC: Yeah, I think, when I think of IVR, I think of call center and I think of the objective on both the person that's interacting with the call center and the company that's making it available is to keep the interaction as short as possible. It sounds like the technology is allowing us to go the opposite direction.

[0:06:38.3] PT: Yeah, I think one of the technical terms in the industry are things like containment, keeping the user contained to the IVR as supposed to transferring to a live agent, he's a live agent, it's more expensive and not at just expense, it's actually hard to staff them, even if you have all the resources in the world, it's hard to hire up call centers that are big enough for some companies to even handle a track even with unlimited budget.

Containment's a big one, another one's first contact revolutions, modular KPI in the industry so that you can have a – you don't have to call back, it resolves within that first call. It's not always about short but it tends to be about can you build self-service, being able to let the user call and interact with the system that they can solve their problem automatically.

I'll talk about that in my talk a little bit about some of the statistics that have come out recently showing that today, especially with younger generations and consumers, really don't care whether they're talking to a machine or a human, they just want to get their problems solved and in some cases, there's even some who prefer to talk to a human, you know Sensitive – sorry prefer to talk to a bot, maybe a sense of situation, you know? You have to call in and talk about I need a change of flight because we have a death in the family or something.

In some cases like that, it's sensitive, you don't actually want to talk to a person even though people have this feeling of emotional intelligence is so important. Sometimes it's actually, you don't want to get into it, you just want to get your transaction done.

[0:07:59.4] SC: Yeah, you mentioned your talk, what's the title of your talk?

[0:08:04.5] PT: It was a long title, this Critical Factors In Conversational Interfaces, Design and Commercial Interfaces, I have the whole thing memorized.

[0:08:13.7] SC: Kind of lessons learned and best practices that you've come across along the way.

[0:08:17.3] PT: Yeah, it will be a mix of lessons learned and best practices as well as some of the newer things we've discovered and developed at Nuance.

[0:08:22.6] SC: okay, walk us through those, sounds like a fascinating topic.

[0:08:27.4] PT: putting me on the spot, I don't think I have them all memorized at this point but –

[0:08:30.4] SC: Well, practice your talk.

[0:08:32.3] PT: I've got about I think 10 at this point that we're going to talk about? I think I had previous in this talk where I had six. That talk was 20 minutes and so I thought, okay, 35 minutes, I've got to beef it up a little bit, yeah.

Well, just expand upon some of the ones that I briefed anyone through. I kind of go, there's like a high level of things like context and personalization. If you're on a website and you're browsing,

We do a lot with banks and insurance and those kinds of enterprises and you're in the auto insurance part of the website and a chat bot pops up and says, "Can I help you?"

You know, that's the typical thing they're going to say, right? Can I help you? At this point in the website, basic UX, you know, will let you know, you should say, "Hi, so and so, hi Paul, how can I help you with auto insurance?" Those basic principles of UX have not totally been carried over yet into chap out world.

It's those kind of things like building those integrations into the systems, the websites and stuff, a lot of times these chat bots come from a different platform. You have to actually build those integrations between the company's website and the chat bot.

Those kind of thing's basic UX but they're not totally there yet in the chat bot world.

[0:09:41.3] SC: It's funny you mentioned that, I've been kind of on the DL, evangelizing this idea that a lot of you know, we've learned so much about user experience design in the web world and in mobile and kind of in these other interface technologies that hasn't really made it into or we haven't formalized to a degree yet or in and around AI and artificially intelligent user interfaces.

I think the – I called it intelligent design but that's kind of an overwhelming term. I haven't come up with the ideal replacement for that one but as certainly, you know, chat bots are the kind of the tip of the spear so to speak but even devices like your nest thermostat or something like that, there's – I think there are like unique things that you need to take into account to make sure that users are comfortable with the fact that this thing has some intelligence and also kind of signal to them how to interact with the intelligence.

[0:10:45.7] PT: It's interesting now, Nuance has been doing, we call VUI design for a long time or voice user interface design. These are people who design, many of them backgrounds and linguistics page, decent linguistics even who design the flow of the conversation and will do, you know, user testing AB testing, et cetera, to figure out what the rest flows are and these people with this experience are being scooped up now by Google and Amazon.

You look at the people now, Google has a whole VUI design team now for a system as Amazon, a lot of former Nuance people because we've been doing this for a long time in the IVR world

and it turns out like a lot of those the basic principles of how a conversation works without a screen, transfer over to these kind of IOT devices.

It's different when you have a screen, when you have a screen, we actually know a lot about web design UX and that's another problem, right? There's this area of overlap between a thing of a VEN diagram of like conversational user interfaces, voice interface design, web user interface design and that little spot in the middle of the overlap in the VEN diagram where this weird world of okay, well the conversational principles aren't going to cover it completely.

The web principles are going to cover it completely, you have to kind of merge those two together and figure out things like "okay, well, on this part of the website, the person's browsing, looking at whatever product, the conversational agent needs to know about that stuff."

That's one of the things I'll be talking about, another thing is security, making it easy for either authenticate by electrician password or SSO as well as nuance has voice biometric products so he could easily transfer you over to say like my voice is my password, use a voice print to identify.

[0:12:21.6] SC: I don't get the sense that that's very popular.

[0:12:24.7] PT: not yet, no. It's really new, really cutting edge. It's easier in mobile devices when the mobile devices where we're building virtual systems there where the whole interface has spoken, it's a little more part of the flow but prototype systems now where you can actually voice printed NFI for a bank on the web.

[0:12:43.2] SC: Wow.

[0:12:44.4] PT: Yeah, that's a huge friction point. I think of the places that recognize my number and based on my number, you know, associate me with a record like the airlines for example have done a really good job of coming up to the speed on this in the past few years.

Then there are other places where they're like, "Okay, type or say your password" and like, well, you know, it's like 26 characters and it's in my password save thing or there's no way I'm either going to type it or say it.

[0:13:13.1] SC: Yeah, this is an art, it's a big area of AI machine learning, you know? These systems are all AI machine learning driven, these voice biometric systems, there's also face printed notification now, we do some work there and we'll start to see other kinds of biometrics too like behavioral biometrics, they've heard of this where the way a person interacts with a website, the way they move their mouse.

This is the cadence in which they type et cetera. That also creates a unique fingerprint of that person very hard to spoof, finger print. I think my first experience with that is with Coursera, like when you taken a Coursera course like the Andrew Ang, you know, deep learning course like you, there's an honor code and you have to type out this honor code thing that says that this is you and when you're taking one of their exams, you type this passage out again and then they use that to verify that you're you.

[0:14:00.3] PT: I haven't seen that, that's cool.

[0:14:01.1] SC: I think if you don't, if it doesn't think you're you. You have to take a picture with your webcam.

[0:14:04.7] PT: Wow, okay, that's really cool. Yeah, that's another yet, so reducing that friction point and I'll be spending a while talking about some new techniques that we're using to try to incorporate unsupervised learning into our pipeline. This is kind of a frontier area for AI right now, the majority of it is all supervised at this point but we're looking at methods where you can take in a set of chat logs or voice conversation transcripts.

The big thing with chat bots and with conversational systems is the first layer is the NLU the intent classification, what's the intent of the partner was saying, such that then you can then respond to that intent, whether it's like check my balance or pay a bill.

We're working on looking at data sets and extracting those intense automatically through unsupervised learning.

[0:14:56.5] SC: Kind of double click on that and give a little bit more detail there?

[0:14:58.8] PT: Yeah, sure. Unsupervised methods, one of the big things they tend to do is group things together automatically. Yeah, clustering or hierarchal clustering or various kinds of methods that bucket things together. I can't give out all the secret sauce there but that's part of it, yeah.

Part of it involves that and part of it involves other steps that can actually identify what the intents are of those cluster automatically. As a first pass, this used have to all be done manually, we call them speech practice and also our data scientists will have to go through the data to figure out what the intents were in a large data site.

As well as interviewing subject matter experts at a company, now is the first pass but today we've managed to cut down a process that you know, used to take hundreds of hours down to a few days, a month down to a few days by using this new practice that we call intent discovery, whereby you can bucket the data and then automatically identify what the intents are in the data.

Then, you can use that bucketing to automatically label the data so you get a first pass at like a label data set and bootstrap a model or train a model off of that. Put them all on a long line. Now, the idea here is that the system won't necessarily have the same accuracy as it would with a hand tuned system.

We've put hand to 85, 90% accurate, one of these, bootstraps might be like 65, 70% accurate. But then what we do is we put that system online and for all the questions that the VA or the virtual system or the bot doesn't know, we can pass off for one turn to what we call hidden agent, a person, a human loop, they can then check what the intent was for that and send it back in a system.

The user ends up with this seamless experience of they're just talking to a bot and be with like a five second delay when it goes to the person or 10 second delay where it says, "Hold on, I'm checking with one of my partners." Then we can use that data then train in the future and then we won't have to have that missing data, that loop in the future. That makes sense?

[0:16:52.9] SC: It makes a ton of sense, I'm wondering, does this create a new business model for nuance where in previously I'd imagine you selling an enterprise, some set of technology where as now, it's the technology but also the service, are you providing the agents that does this or are you providing a console that they can have their own virtual agents doing the teching. Or both?

[0:17:17.3] PT: Yeah, both. Yeah, a few years ago, was it about a year, two years ago, we acquired a company called Touch Commerce which is a live chat platform, they provide a whole suite of live chat, they can provide the agents to you or they can just provide the interface to you.

We have that which is really cool too because now we can – if a virtual assistant doesn't know the answer or chat bot doesn't know the answer, we could, if a company wants to partner with us to transfer that way, transfer direction to our live chat but we also talk with lots of companies who have their own live chat platforms.

A big one is Sale Force, they provide live chats with really popular one these days, transfer to those agents, we have a console we can provide to allow companies to have this – to use their own pool of live agents to do this human assisted step or we can provide the software and the agents, all different ways of working on it

We can customize and all different ways with that, not to be too salesy here but you know, we have all different ways of working on it and yeah, it is kind of a new line of business and a new way of thinking about it.

[0:18:16.7] SC: Okay, all right, what else?

[0:18:19.1] PT: Targeting, that's another big area, When do you have a chat window actually pop up? You know, I think –

[0:18:27.4] SC: Based on the way it's implemented now, always.

[0:18:28.9] PT: Always, right? Yeah. It seems like when I use these systems, I just didn't have any rhyme or reason for, I feel like a lot of times, these systems, it's just a clock or something, right? If you've been on the page longer than 30 seconds, it pops up.

[0:18:42.4] SC: Or if you're visiting the site on your mobile phone, do it, you know, immediately, obscure everything else out there.

[0:18:48.5] PT: Yeah. See, there's the UX issue it totally worked out. Yeah. We have a targeting engine that's constantly undergoing new prototypes and stuff but that's aimed at like why don't you interact with the user, how do you interact with the user and what way, what language do you use even down to the colors and stuff on the interface and that targeting engine is based on all kinds of things like how long has the user been on the page and certainly one of them but what have they put in their shopping cart. Like they put something in their shopping cart and walked away.

So there is all of these different inputs and systems that you can use to then tune when you're going to pop up that trap bot to ask if they want to talk to an agent or talk to or Nina is the name of our registrar, talk to Nina, talk to a chat bot.

[0:19:33.7] SC: And assuming it's machine learning driving that ultimate target decision, how transferable are the models from one customer or one website to another? Are you training this models on a customer by customer basis or is there, you know do you have industry models or is there a generic model?

[0:19:55.9] PT: Yes, right.

[0:19:56.5] SC: That out performs just show the jet bottle.

[0:19:58.8] PT: Yeah so there are. We do have generic set ups for various verticals but what it comes down to is a lot of this is this is a very difficult subject right now these days for us because a lot of our customers really feel very strongly. They don't want the data shared, they don't even want the models shared you know?

[0:20:16.8] SC: So even when their chat bot pops up on their website?

[0:20:22.1] PT: yeah it's a very tricky area that I think that our company, a lot of companies today are having to deal with. Which is like how do you both be a competitive AI machine learning company while at the same time protecting the data of your customers in a way that they're comfortable and they're doing all the compliance that you choose. It takes a lot of our customers who are banks so they even be legally you can't unless there is actually principle that consent. For example, we deal with the government agency in Australia and I was reading one of their privacy policy relating to data the other day.

And they were saying unless you have explicit consent from a user like a policy consent to use this or you sign some check box or something, they can't use that data for anything else. Tuning a model even. So this is really I think, we talk about all the exciting stuff in AI today and all the amazing things that are happening, what it comes down to for a lot of businesses, these are the real problem today. It's not how do I scale a deep learning model or how do I productionize this system and get it working on going from one cluster to a 1,000 computers or whatever, it's the legal problems, how do you actually deal with the data in a way that is safe for the company and for the users as you know, really with a huge summing lot.

[0:21:41.0] SC: Great, yeah I was just reading, I forget what, I think it was Argo or someone else's newsletter or something like that that was talking about how there's this huge grey area around copyright and data sets and basically everyone who is using for the most part these public data sets is kind of flying under the radar but there's a potential exposure where there's no established president around the extent to which copyright on a data set flows into the model for example.

And so potentially someone using this copyright data set to train the model could be creating a model that has potentially owned or has some copyright liability with someone else. So that's an

example of this kind of meta concern that your customers are thinking about. Do you have any perspectives on the rise of the consumer voice interface devices? The virtual assistance, your Alexas and your Google assistants and things like that?

[0:22:48.5] PT: Yeah, it's interesting. I think a lot of that stuff that has to do with their API's with those kinds of systems, right now we have our Chapa integrates with Alexa, it integrates with Google home but there are certain things that we can do and there are certain things we can't do. So one thing we would really like to do, a lot of our customers like I said are banks right? And with banks this voice biometrics has become very popular recently.

But one thing we can't do right now is voice biometrics over Google and Alexa because they only send you text. So at this point, they don't actually send you the audio signal. They do all the ASR for you, speech recognition and the TTS for you, the text to speech and they have no way of sending out those signals. So I think this is something that we had huge strides in terms of the openness and their compatibilities platforms with Alexa and now Google home.

But for years this was a big problem with Siri right? With Apple, they had no ability to do third party integrations. You couldn't say to Siri like, "play a song on Spotify for me." So this is definitely a big topic for Nuance right now. This idea of what we call it cognitive arbitration where you have agents that sit like Uber agents that sit in the middle of all of these different IOT systems and coordinate those for you and bring that data together.

Bring those systems together and those API's together to talk to these systems. I know this is also a vision of Amazon as well where being able to sit and have their agent that goes out and have all of these different skills that they come to. So this is a newer topic for Nuance that we were working on more recently is how you can use intelligence and reasoning and other kinds of machine learning to build agents that can sit in between all of these different IOT devices and talk to each other.

[0:24:32.1] SC: Is there a role or any emergence of like standards or standardized approaches like for example, there's tons of work that's been done around federated identity and the web and now this is, you are introducing a whole other layer around biometrics. Does any of that kind of work transfer here to do you think or we're just too early for that?

[0:24:56.1] PT: No, I think it will. I think this is we're in the very early days here. So it is basically like I don't think the very beginning of the internet when these standards were being built out like what's HTML and how does it work. That's where we are today with AI so this is usually the thing like you were up with the data, those kinds of issues have been solved like yeah, how do you? There's no Apache license for data at this point.

Whereas that stuff to season software engineers and pragmatics down, it's like second language. They know what licenses mean, they know which ones they can reuse, what code can be leveraged which camp, why you have to declare, what you don't, etcetera but that's the person exist for data bits. An interesting point though I hadn't thought about it yet but yeah, I think that will come about and similarly with these standards around their operability.

I think that's something that we have teams at Nuance we're working on and trying to reach out to different partners because this is not a problem that is going to be solved by any particular company just in the same way that Jayson's standards can't be solved by Google. They can't just make a really great Jayson parts and everybody take it and put it in their browsers or put it in their systems. It has to be standards based and community based. So I think we are thinking about these kind of inoperability standard to this point, very early days though.

[0:26:07.0] SC: Great, any other things that you've covered in your talk that you want to share with us?

[0:26:13.4] PT: Yeah, I covered most of it. There is a couple of things in the front matter that I've covered at the talk that I was talking about which is that in this world of chat bots and virtual assistants, it's surprising by how much of it isn't AI and isn't machine learning and we really, there is a lot of companies that will have a dot AI in their name but everything is still based on regular expressions and rules and that sort of stuff.

[0:26:39.0] SC: Or people.

[0:26:40.8] PT: Or people, yeah. People big time, it's not always necessarily a bad thing but it's often part of the company's strategy in terms of having everything in the beginning be run by

humans as a way to gather data and then overtime learn from it but that is something we try to help educate our customers about is how much of this actually is AI and machine learning and how much of it isn't. Today most of the AI and machine learning for us happens on the language understanding side of things.

So when the input comes in being able to categorize it using statistical models and OP models in order to route to the right response. You know to get the right response to the question. We can also do things like inner deep extraction, extract the cons of yourself saying something like, "I would like to order a pizza" or "I'd like to order a large pizza, a large vegetarian pizza with peppers and onions." I almost said pepperoni.

Be able to identify both that the intent was to order a large pizza that the toppings and the size and that sort of stuff fill out those categories that kind of technology. Now some of that stuff is standard but there is still quite a few different platforms out there for doing chat bots that don't offer that kind of – that level of machine learning NLP but beyond that –

[0:27:55.1] SC: I've played with that stuff in the past like API.AI, it's a very manual and tedious process to build out those entity trees and so it sounds like what you are describing is a way to learn some of that from the actual data itself.

[0:28:10.7] PT: Yeah exactly. So when it comes to our intent discovery and bootstrapping process, learning the intense automatically and building out – you could call it an ontology. The group of the intense and the various concepts that are associated with those intents but yeah, we try to just encourage people to be discerning and try to figure these things out when you are looking at a platform. Today I think the next frontier is going to be on the other side though.

There really isn't any system on the market today including ours that can automatically learn how to answer the questions especially if they are complicated back and forth dialogues. So if you acquires like a back and forth conversation, those tend to be built out manually either as enterprises – when the enterprise you know by talking to subject matter experts or people are doing it themselves by figuring it out themselves that way.

That I think is the next frontier for this technology is learning how to answer questions and dialogue there's a lot of it.

[0:29:08.4] SC: Now there are tons of people who say they can do that.

[0:29:10.9] PT: Yeah, there's a lot of them. What they tend to do though is you have to feed the system a question and answer pair then it can learn to map new questions to those answers, right? So this is question and answering but it is not dialogue. If something exists sort of an FAQ can you identify if the question takes a different format that will actually map to this question that is in the FAQ palatable format meaning it could not even be a FAQ.

It could be like a list of a thousand question and answers pairs or whatever but I mean cutting edge research, I don't know if you are familiar with the Stanford squad dataset.

[0:29:45.4] SC: Q&A database.

[0:29:46.4] PT: Yeah, Q&A dataset where it's trying to answer questions on Wiki PDR. This is pretty cutting edge research at this point. There is research teams across the world competing in this kind of thing but even that still is focused on you know the answer is in the article, you know? We are talking the kinds of Nuance's working on today that is one thing I forgot to mention. We got a product we're working on right now called Nina Knowledge.

Where you can basically push a button to ingest a website or a set of documents and then start doing push and you are answering on it and it does leverage some of the technology that's being used to answer like those. You need to work with that squad dataset but it also leverages technology for information retrieval, search. So it combines a few different areas, machine learning as well as NLP and question and answering and I like to think of this as low hanging fruit really.

Because these kinds of that you actually have just a one shot answer to, those are the candidates for automation today. The things that we still really are still very early days in the research in is how do you actually learn a back and forth conversation that requires multiple

questions and feedback onto your conversation with somebody from data. I think it's the next frontier.

[0:30:56.0] SC: In the case of Nina Knowledge, to what degree is it you're doing like transfer learning off a model trained on squad ad applying that model to the website that is being ingested or is this process including like training up a new model on that website?

[0:31:15.4] PT: Yeah, the process is strictly we start from scratch on each one each time. We haven't gotten transfer learning or anything like that into production yet today but those are certainly areas of research we are working on. So one area where I don't know if you'd caught transfer learning or not but you can certainly do things like learning word vectors from one data center and apply it to another. That is certainly a form of transfer learning.

What most people think of transfer learning I think today they are thinking, "Okay I built this big multi-layer neuro network and I'm extracting a piece of it and then raising it on another dataset." I think it's still early days for that sort of stuff but certainly when it comes to word vectors or other kinds of vectors, paragraph vectors, document vectors etcetera that stuff can be transferred and it can be very helpful in improving the quality of the model pretty quickly.

[0:32:00.5] SC: Where are we in terms of, we've talked a lot about identifying the intent of an auto rent or something that is typed into a chat bot but where are we in terms of more realistic kind of dialogues that have more multiple intents or hidden intents or things like that and this is maybe a slightly different direction for their question but for a while, we have talked about the IVR systems being able to identify the emotion or change or escalate the way the call is handled based on the emotion.

You know when I am frustrated I try to make sure the IVR knows that I am frustrated and it never makes a difference you know what I'm saying?

[0:32:41.0] PT: I really love this question particularly the emotion part of it. I really love that question because we've been working on it a lot lately of what do you do with tax incentivement right? So that's basically 10th of being in the seminar although there are some models now that

classified things into the five base emotions of anger, happiness, frustration, sadness. It's these old models of this. There are some models that claim to be able to do that.

I don't know how accurate they are but sediment that is at least one that we're pretty accurate. We can get pretty accurate on that for the most part. At least global sentiment of a sentence. That isn't always sufficient because there can often be two sentiments in a sentence here, I love this but I hate this what do we do there? So getting off on a tangent but the reason I think this is really fascinating is because the real question is, "Okay great so what do you do with it?"

What do you do when you know someone is pissed off? What do you do when you know someone is frustrated? And that's a really hard question because it tends to be what a company, what a customer wants to do is figure out a way, it's about containment. So how do we handle this without escalating to a person because escalating to a person is expensive but when you have a frustrated person, the only thing we know how to deal today.

The only thing we tend to do today is escalate to a person. So I think what it's going to come down to is understanding that these aren't black and white things that there's – just like we have confidence scores and probabilistic models that give you ingredient of how sort have confidence and say a right answer. There is going to be ingredient in terms of sentiment. So if a person is saying something like, "You know I am having trouble with this" then you might be able to say:

"Oh okay I'm sorry you are having trouble with this and I really wish you were doing a better job there but can I offer you this tool, this trouble shooting step" or this trouble shooting dialogue. Whereas if the person is really upset then yeah, you probably just want to say or you might be able say, "Would you like me to have a technical support person to contact you directly? You can just hold on and we'll call you or do you want to proceed with our trouble shooting online?"

So I think although we still aren't there a 100% with recognizing these things, the question is what do you do with it once you have recognized it? It might be even a harder problem to deal with you know? Because great, we can get like – we end up with a 100% or 99% accuracy of knowing how angry someone is but how does that help you, what are you doing?

[0:34:57.7] SC: But your customer does not want to escalate all of those to their most senior rep or manager or something like that.

[0:35:02.7] PT: Yeah exactly. Right, so how do you handle it? What do you do with it for enterprising? How do you actually use that data in a way that is interesting? Again this doesn't actually come down to the eye problem, it comes down to the user experience problem or design problem.

[0:35:14.9] SC: Yeah, I wonder if there is like company is all over of logged on to net promoter score is a way to measure customer satisfaction and I wonder if there is a net IVR anger score or something like that.

[0:35:26.6] PT: Yeah.

[0:35:27.2] SC: You guys should do that.

[0:35:30.6] PT: Yeah, our customers ask about this stuff a lot. So it's always on front of people's minds but you do today when you go down that dialogue when you go down that path it's, "Okay so then what?"

[0:35:44.0] SC: Yeah, awesome. Well Paul, thank you so much for taking the time to sit down with me. I am looking forward to catching pieces of your talk and how are there ways that folks can find out about your research or connect with you on Twitter or anything like that?

[0:36:00.5] PT: Yep, my Twitter handle is my name with the first initials switched. So @taulpepper on Twitter. That's probably the best way to get at me these days.

[0:36:11.7] SC: Okay, nice. Well good luck with your talk tomorrow.

[0:36:14.7] PT: Thank you.

[END OF INTERVIEW]

[0:36:20.5] SC: All right everyone, that's our show for today. Thanks so much for listening and of course, for your ongoing feedback and support. For more information on Paul and any of the other topics covered in this episode, head on over to twimlai.com/talk/52. For the rest of this series head over to twimlai.com/aisf2017 and please, please, please send us any questions or comments that you may have for us or our guests via Twitter @twimlai or @samcharrington or leave a comment on the show notes page.

There are a ton of great conferences coming up to the end of year. To stay up to date on which events we'll be attending and hopefully to meet us there, check out our new events page at twimlai.com/events. Thanks again for listening and catch you next time.

[END]

