EPISODE 78

[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.

This week on the podcasts, we're running a series of shows consisting of conversations with some of the impressive speakers from an event called the AI Summit in New York City. The theme of that event and of this series is AI in the enterprise. I think you'll find this series really interesting, and that it includes a mix of both technical and case study-oriented discussions.

Now, I won't actually be attending the AI Summit this week, because I'm in Long Beach, California attending the NIPS Conference. There are a bunch of TWiML listeners here, and I'm hoping to meet as many of you as possible. Yes, I have stickers.

If you're here at NIPS and you're actually listening to podcasts this week, please reach out to me either via the event app, the NIPS event app where there is a TWiML listeners thread, or via Twitter where my handle is @samcharrington.

Before we proceed, let's quickly talk about the podcast schedule through the end of the year. If you subscribe to my newsletter you know that I've been on the road for a couple of weeks now. After this week's series, we've got two more series coming before we break for the year with our last show running on December 22nd.

Now if you're lamenting two weeks without your favorite Machine Learning and AI Podcast, trust me with these two series, the first from the Amazon web services reinvent conference and the next one from NIPS, you will have plenty of great content to tie you over until we get started again on January 8th.

Thanks to you, 2017 was a great year for the podcast and we plan to close it out strong. Keep your ears open the next few weeks and we hope to hear from you.

Please note that on Wednesday, December 13th we'll be holding our last TWiML online meetup of the year. Bring your thoughts on the top machine learning and AI stories of 2017 for our discussion segment.

For our main presentation, Bruno Goncalvez will be presenting the paper Understanding Deep Learning Requires Rethinking Generalization by Chiyuan Zhang from MIT and Google Brain and others. You can find more details and register for the meetup at twimlai.com/meetup.

This AI Summit series is brought to you by our friends at IBM Power Systems. IBM Power Systems offer server's design for mission critical applications and emerging workloads, including artificial intelligence, machine learning, deep learning, advanced analytics and high-performance computing.

IBM Power Systems benefit from a wide range of open technologies, many stemming from collaboration with fellow Open Power Foundation members and their design to deliver performance efficiently, whether deployed in private, public, or hybrid clouds.

To learn more about the IBM Power System AC922 platform for enterprise AI, visit twimlai.com/ibmpower.

Continuing with last show's theme of platforms for enterprise AI, my guest this time around is Mazin Gilbert, Vice President of Advanced Technology and Architecture with AT&T. Mazin and I have a really interesting discussion on what's required to scale AI in the enterprise.

You'll learn about a new open source project that AT&T is working on to allow any enterprise to be able to do this. You already know by now that I geek out when it comes to talking about the intersection and machine learning and cloud computing and this conversation is no exception.

Be sure to let us know what you think by posting your comments or questions to the show notes page at twimlai.com/talk/78.

Now, on to the show.

[INTERVIEW]

[0:04:17.6] SM: All right, everyone. I am on the line with Mazin Gilbert. Mazin is Vice President of Advanced Technology and Architecture with AT&T. Mazin, welcome to This Week in Machine Learning and AI.

[0:04:29.9] MG: Well, thank you very much for having me today.

[0:04:34.0] SM: Yeah, I'm excited to jump into this conversation. As is the tradition here, I like to get started by having you tell us a little bit about your background and how you came to get involved in AI projects.

[0:04:45.3] MG: Thanks. I started looking at AI when I was just doing my graduate degree in the University of Liverpool. I got fascinated in the mid-80s about how you could use these AI technologies to really teach machines and computers. Then at the time, computers and machines were very primitive. How you can get these machines and computers to do what humans can do. They can think and say and listen the way humans are. I thought that was

extremely fascinating. It was an area that's caught my passion. That started my graduate degree and I've been involved in AI machine learning ever since.

[0:05:30.8] SM: Wow. What did you do your graduate degree in?

[0:05:33.8] MG: My graduate degree, my PhD was on artificial neural networks for computer speech. Basically at the time, the problem I was trying to solve was that how do you get computers to speak like humans using artificial neural networks.

[0:05:52.1] SM: Okay. What did speak like humans mean for you in that context?

[0:05:55.7] MG: It was really speak like humans as the way that you and I are speaking today. The goal was to pass the touring test. Is that how can you get a machine to speak to you for so many seconds or half a minute without you knowing that was really a machine?

Clearly in the past 30 years, a lot have changed. But that was at the time, computer speech was extremely very machine-like. There was a lot of promise a the time of artificial neural networks, could be the new technology that can completely change that, because of the association of artificial neural network with the real neural networks in the brain.

[0:06:39.5] SM: That's turned out to be true in terms of all of the advances that are happening now with deep learning.

[0:06:47.0] MG: Yeah, absolutely. I mean, it's quite astronomical to see how much deep learning, which is really based on artificial neural network, how we have managed in the past 30 years to use those in a lot of different areas beyond just computer speech. They're very powerful tools. I think what – a lot had changed in the past 30 years, but the basic algorithm, the basic technologies from the 70s and 80s are still in use today.

[0:07:18.9] SM: Did you go directly from grad school to AT&T?

[0:07:22.8] MG: I did. I actually came to visit AT&T Bell Labs in the late 80s while I was doing this work. I was really fascinated by the place. It was like a kid in a candy store. It's like you get to do your passion and your hobby and get paid for it too. It was really an amazing experience. I knew at the time when I met a bunch of the folks that I used to sort of know them, when I used to write these papers, they used to be like a reference in my book, or reference in my papers.

Now I get to see them face-to-face and some of the pioneers that invented a lot of amazing technology in digital transformation and transistors and others. It was an amazing experience and it was very clear that was the place for me. Honestly, as soon as – I couldn't even wait to finish my graduate degree. In fact, I ended up doing – I finished my graduate degree at AT&T

Bell Labs, because it was an area they were very interested in and neural networks. That's how it started.

[0:08:30.2] SM: Wow. Can you give me – certainly AT&T and the labs are doing a ton of things with regards to AI. How can you give us a landscape of the different areas that you're working in?

[0:08:48.2] MG: Yeah. AI is a very critical area for us and has been really for decades. The way we think of AI, it's a way of doing data powered smart automation. When you think about it as just data-powered smart automation, it really touches every part of our business. The machine learning side is what makes it data-powered, and the AI is what provides this smart automation. That's how we think of the two and their relationship with each other.

The way we look at this is that where are we spending a lot of our dollars today, and that we need something, new capabilities that are disruptive in the market, disruptive for AT&T really to try to break down our cost structure and provide better experience to our customers and better products to our customers. We've identified a number of what we called major pillars several years back. A lot of our activities are really focused on those pillars.

[0:09:49.9] SM: What are they?

[0:09:51.1] MG: The pillars are number one, is infrastructure and engineering. Think of that as a company like AT&T that invests significantly in capital and infrastructure in building 4G and 5G and we're in the homes and we're in your devices and everything. There's a huge infrastructure that we put in place.

We spend a lot of time and capabilities in applying machine learning and AI to really do smarter investment and spending of building out our infrastructure. Where do we put the small sales? Where do we put the macro sales? Where do we invest more in what part of the US where there is most traffic? Or we think there will be traffic over the next year to three years. That's infrastructure engineering.

Number two is in the area of customer experience. I think customer care and customer experience is the traditional approach, as you know the whole industry went into that. With AI one day when they started using speech recognition and speech synthesis and some natural language processing to start doing part of what humans would do today, at least some of the basic things of let's say understanding the intent of the customer and routing the customer. We're doing a lot of that and have been doing that for chat, for voice, for internet, for e-mail. Basically, the customer experience side of the game.

The third pillar that we also invest a lot of time is obviously security. When you look at the size of AT&T, and we provide our customers with a secured network both consumer and enterprise.

Those hackers are becoming smarter and smarter. That we have to apply the best and the latest technologies to identify even trends or patterns of an initial attack to any of our customers, or to our network, so we can address that quickly before it becomes a big issue. There are millions of those that are attempted every hour of the day. That's a third area.

The fourth area is our network and operation. As you know, we run one of the most complex networks in the world, and that network has evolved over the past century since the invention of the telephone. You can imagine that there is a lot of push to try to make it zero touch, to make it completely automated, to make it completely software-defined. So there's a lot of work we do in that area.

The last one I would mention is the area where we look for a new revenue opportunity. We apply AI and machine learning for advertising, media analytics. In the future, a lot of our enterprise businesses that they sell to customers, whether it has to do with the internet of things, whether it has to do with smart cities, healthcare, the opportunity space is big. AT&T is a very diverse company. As you can imagine, there a lot of opportunities there.

[0:13:01.1] SM: That is a ton of stuff. In your role, what's your specific organization's responsibility with regard to these five pillars? Is it staying ahead where this technology is going and then feeding it back to businesses, or are you building our projects and products that these businesses can take advantage of? How do you think of the role of your business?

[0:13:30.6] MG: It's along the line of what you said. I lead an organization, the advanced technology and architecture. This is the research arm also for AT&T, and part of my responsibility is to ensure AT&T is always one step ahead and two steps ahead.

It's not just understanding the technology, but also understanding where are the opportunities and the applications where we could provide value at our customers and better products and services to our customers. That is part of my organization, part of the job that we do. But beyond that, it's also about – one thing I think you're just striking about all the pillars I mentioned, we're not a company trying to build that application for AI. We're not trying to build a voice personal assistant and make it available to our customers. We're not.

The use of AI and machine learning for us is across many different services, many different opportunities where we spend billions of dollars today in each of these pillars. There are not just 10 or 20, but they're literally in the thousands of applications where AI and machine learning is required. Part of my responsibility is to make sure that we can scale, and that each of these application is not yet another service, new technology, new operation of its own.

You have to provide some harmonization to make sure that all of these thousands of applications when we get there, they all are built on a common platform, a common technology,

common capability. So as technology evolve, you are benefitting everybody. You're not building just a silo application for one set of users.

[0:15:21.8] SM: One of the challenges in building platforms when we're so early in the technology cycle is that it's not clear what the pieces of that platform need to be. How are you addressing that and other challenges that you run into given the scale that you described and the fact that you're trying to apply AI very broadly across these thousands of opportunities?

[0:15:49.1] MG: That's really a great question. You're absolutely right. I think that, although the technology behind machine learning and AI have been brewing for decades, but we are just at the tip of the mountain touching that sort of – our initial – we're getting just a flavor of well, where are all the opportunities? Until you understand all the opportunities, it's hard to say, "Well, here is the platform that's going to set forth all these opportunities." You're absolutely, absolutely right.

The way we're looking at this is a little different. We're looking at this coming in saying, "Well, you are absolutely right and we have to enable – the technology is going to continue to change. There's going to be new applications, there's going to be new tools, there's going to be new capabilities. The way we need to think about the platform, we have to build an open platform.

It's got to be an open platform that enable rapid innovation, not just within AT&T, but within the global ecosystem. It's a platform in which that our vendors and our suppliers and our partners are part of the development of this platform, part of the ecosystem. It's we are all sort of joining hands together to building, harmonizing the core foundation behind AI across the world. It's because every company out there have the same challenge, the same issue, and every company is going to end up doing their own thing, unless we really build the foundation platform to be open and transparent.

[0:17:29.0] SM: I came across reference on your website to Acumos. Is that a platform, or the platform that you're describing?

[0:17:40.5] MG: That is absolutely right. How this started is that when – about a year ago when we were looking at the applications, AT&T has deployed or about to deploy an AI and machine learning. There were tens of – at least, the ones that I have tracked, I am sure there are plenty more, the one I have managed to track, they're tens of applications already deployed, or about to be deployed.

When I looked at those, it's sort of every application supported by every business unit is different. It's different technology, different vendor, different tools, different operational support, and it was very clear that it just – if this is the direction we're going as a company and we absolutely are, it's just not possible to scale. It's straightforward to build one, or to build two. But

to go and build now and scale 20 and take 20 to 2,000, you can't. It's just not possible without having some commonality.

Acumos was really born as a result of that, thinking of AT&T as a customer, not as a builder of the platform. It was really built to try to solve three key problems that AT&T was being challenged. We just discovered since then is that pretty much most companies are being challenged by those three problems.

The first problem is there are a lot of tools out there for machine learning and AI. Wait another week and somebody else is going to come out with another tool. We love some of these tools. Some of them are very specialized. In some market, some of them are specialized by the company. They're great tools. There is no one-size-fits-all.

The first problem we're wanting to solve even with an AT&T, we found out many people are using many tools. First, we try to force everybody to use one tool. We actually put out two, three years ago an open source tool for machine learning and a statistical modeling called RCloud. We were hoping to force everybody in AT&T to use it, and frankly we failed. We ended up finding some people use it, and some people like to use other tools that's not even built by AT&T.

We moved away from that approach and we decided that the problem we need to solve with this Acumos; number one, Acumos needs to harmonize across all the tools. If I use tool A and you use tool B, even though we may not be working with each other, we may not even talking to each other, but the output of my tool A and the output of your tool B need to be Legos and not snowflakes. These things have to connect. These things have to inter-operate with each other, not being sort of every tool is its own silo. That's number one thing that Acumos is attempting to solve and that's the problem which we're going to solve there. Harmonization.

The second problem we were trying to solve is that – I think you know this very well, is that to build machine learning models or algorithms, it's very difficult. It's not easy to do. People talk about hiring PhDs and people with graduate degree. Well, how many of them are there today, and how many companies require that skillset?

Everyone has been talking about lowering the barrier, making it easier. But you know what? It's never has been easier. It's difficult. It's difficult to do. It requires some extensive understanding of the data. It requires extensive cleansing of the data. It requires you learning about some deep technologies, understanding what an ROC curve, what a type 1 error, what a type 2 error is. It allows you to understand what is a gradient descent in neural network is.

These are very deep things. Unless you go and take some deep courses and play with your favorite tool, whether it's scikit-learn, or Tensorflow, or Caffe or whatever, you're not going to learn this. That's what they do in graduate degrees.

The second problem we wanted to solve with Acumos was that we need to lower the barrier to entry, so that Mazin goes and build a model, and maybe Mazin takes weeks, uses GPUs and takes weeks to build out his model. But then, if Sam wants to go and use that model, it should be available in the marketplace. You should just grab and use and plug into somewhere.

The second thing Acumos provides is a marketplace. Is a marketplace that says, "Think about the app store." I can go somewhere, I can see what everybody else has built worldwide, and I could look at the review, the rating and I can pick what I want. I don't have to rebuild it. We don't have to rebuild everything from scratch. It's very expensive to rebuild. We have to be thinking about reuse and improve as opposed to rebuild.

That was the Acumos marketplace. But what we do in the Acumos marketplace is more than an app store. What is does also, it says, "If I build something and Tala build something else, then her capability and mine are microservices that we can just join together." We can actually put those two together like Legos. You can come in and connect what she has and what I have from the marketplace, create a new application that didn't exist before. These applications, these microservices are interoperable.

The third thing we wanted to solve is to say, when you build a model with any tool you want, I'm going to be able to deploy this model with any cloud and I want it to be minutes. I don't want to learn what a docker container is. I don't want to learn what Jenkins job is. I don't want to learn what a Jira ticket is and I don't want to learn the internals of a third party cloud. It should be really like, I build it, I submit and it takes me all the way to its running on a third party cloud. That's sort of, again what Acumos is trying to do.

[0:24:06.3] SM: Okay. There is a ton in there to dig in to. I noticed on the Acumos page that there is an affiliation with the Linux Foundation, or at least a project. I'm assuming that the project is being managed under the Linux Foundation. Does that mean that – or taking another step back. It sounds like a lot of the things that you described are almost outside of the realm of what we think about when we think of traditional AI, machine learning tooling.

It's more like the kinds of concerns that platform as a services come along to try to help with. Is the implication of that in the Linux Foundation involvement that this is built on Cloud Foundry, or is taking a similar approach?

[0:24:52.7] MG: Number one, you hit the nail on the head. This is Acumos is not competing with the many tools that's out there today. In fact, we want to encourage communities and

companies to innovate in building the next big better algorithm. How do you do deep learning now with a thousand layers and instead of 20, 30 layers? How do you have trillion of neurons as opposed to a billion of neurons? We absolutely want companies and communities and industries and universities to innovate.

We're trying to fill a gap. A gap that every company in the world, if they want to seriously look at AI and machine learning, they have to fill. Today, that gap is so large that you can hire the best PhDs and you still will never be able to scale deployment adoption and improvement of AI and machine learning services.

Linux Foundation, we have an incredible relationship with the Linux Foundation. We co-created Acumos with a partner of ours Tech Mahindra. We're hosting it under the Linux Foundation. Linux Foundation will host the marketplace and the platform, and we will work with Linux Foundation to grow a community of companies and institutions and universities to really co-create together the foundation of how we want to scale and adopt AI globally.

We are already in discussions with a number of companies, so we don't consider this as an AT&T or a Tech Mahindra effort. We consider that we've started the seed and the Linux Foundation is going to help to grow that seed and really build a lot of beautiful trees with the help of a larger community.

[0:26:47.6] SM: This isn't the first time we've seen things like this. It really reminds me a lot of, for example what Intel was trying to do with their TAP platform. I don't know if you've ever come across that. I'm not even sure they're still working on that. It's a huge problem where you're taking these – trying to integrate these many, many disparate components, like lots of tools.

In their case, I think they also built it on Cloud Foundry and then they had Hadoop and a bunch of data link stuff underneath. These are all very complex enterprise technologies. They're all evolving very rapidly in their own different directions.

I guess my question is maybe pushing back on the whole idea of doing this, like why do you think you can succeed at this. It's such a Herculean task and so many have tried and failed.

[0:27:43.8] MG: It's a great question. Frankly, that if we don't solve this problem, if we don't solve this problem worldwide, the adoption of what you see in the media about AI and machine learning is very limited. It's going to be limited to few companies who can afford to do it, a few companies who have the expertise and a few companies who can build one or two applications here and there.

We have to solve this gap. It is a big gap. You're absolutely right. But I'll tell you what we are doing that's very different. We have experience in this. Five years ago, we started with just as a

big challenge as this one, as Acumos, which is we decided to completely convert the AT&T century, sort of young network that's being developed over many, many decades.

We decided to completely move that to a software defined, using the best technology in SDN and virtualize the network functions. This is like the core of the network. You're now saying, I can build the core of a network like a web company where everything is software running on a cloud environment, commodity cloud hardware. That's just not possible.

Well, we went through that avenue. We started deploying that three years ago. But we realized a year ago, there's something missing in here. What's missing is that we've invested significantly to making that happen. Hundreds of people, a lot of expertise, other companies can't do that, and other operators cannot do that. Just having us doing something and operators are not, that's going to compromise the experience. That's going to also compromise the cost to the end customer. We need all operators, all vendors to be part of this.

We made it open source under the Linux Foundation, and there's something called ONAP. In the short seven months, we've been at this journey this year, we have 50 plus companies, we have 1,400 technical people co-developing from these companies and we cover 55% of the global sub. These are the operators, the China mobiles, the China telecom, the Vodafones, the Iran, Bell Canada and many of the brands that you basically hear about.

We've cracked the code that the way to do is that if we can all chip in, and if we all can build it together, we can create something better than any one company trying it on their own. This is what we're trying to do the same thing with Acumos. This is if AT&T ends up the only one, or AT&T and Tech Mahindra the only one doing this, it will fail. I'll just tell you that flat-faced.

What's going to be different in here, that's not what we're seeing from the companies we're talking to already, I can tell you that right now. What we're seeing is a tremendous momentum. They all realize the big gap. They all realize the big challenging gap that you're talking about. By having a community come together and each of the companies chip in to what they believe is important to them, so we will come in and say, "Well, what's going to work on our native cloud environment? It's got to work on a couple of third part cloud it has to work, and it's got to do X, Y and Z."

Other companies coming in with slightly similar, with different also attitude, bringing it all back together is really what's going to crack this. It's not going to be AT&T, it's going to be a community building it together. All at times that's happened in the past, it's been one company trying to do it, trying to boil a notion. What we're doing in here, there's no one company in the world can do it. It has to be a community, a global community effort.

[0:31:37.0] SM: Interesting. You mentioned ONAP. This is maybe going a little bit deeper down this rabbit hole, but what's the relationship between ONAP, which I hadn't heard of and OpenStack? I'm fairly familiar with AT&T's involvement over the years in OpenStack, in particular working to build NFV deployments on top of OpenStack.

[0:32:00.9] MG: Yeah. The two are really like brothers and sisters. They're complementary and brothers and sisters. OpenStack is the – we are a big user of OpenStack. We've built our AT&T integrated cloud platform based on OpenStack. We take OpenStack. We're a significant part of OpenStack, as part of the board as well. We ingest OpenStack and we build our AT&T cloud. This is what we use as the infrastructure to virtualize our network functions.

When I'm talking about virtualizing the network functions, moving on network to software define, it's all using our cloud and sort of a hybrid cloud, our cloud and third party cloud and OpenStack is a part of that.

ONAP is that once you have that, once you have these network functions and you have your cloud environment, what you need to orchestrate these network functions, you need to automate the design and the testing and the deployment of those network functions, you need to do life cycle management of these network functions. Something could go wrong. They could be a threat. They could be a failure.

That is what ONAP does. It's an orchestration, automation platform for both virtual and physical network functions and life cycle management. The two go together. You need ONAP on top of OpenStack and our AT&T integrated cloud to really try to virtualize these functions.

Acumos and AI is the – a key enabler to enable both the cloud and the ONAP to automate functionalities that are machine learning data-driven. If ONAP comes in and says, "I want to use data-driven capabilities to predict there's going to be a failure in the network in a particular location for a particular customer an hour from now," that's where machine learning Acumos comes in is creating these predictive, sophisticated capabilities to make the ONAP and the integrated cloud data-driven and predictable, and have predictive capabilities.

[0:34:30.5] SM: Actually very keen on digging a little deeper into some of these network and operations use cases, but I do have one question on Acumos before we do that. I noticed, or didn't notice the lack of on the Acumos page and link to like a Github repository or something like that, what stage is Acumos in? Is it a vision, or is there more or something more tangible to it today that folks can get access to?

[0:34:56.7] MG: You always start with a vision, but what it is is not just a vision. If it was a vision, then we wouldn't be having the conversation, I'll be honest with you. It's we believe you have to have something real. But this is a community effort. Like you said, technology is

changing a lot. But there are some basic foundation you have to build, so what we've done is we announce this a couple weeks ago, or whatever, 10 days ago. Next week, literally next week we're moving all the software to the Linux Foundation. We built the software, we built the seat capabilities of everything I talked to you about, the seat software of everything I talked about is being built and we're moving it to Linux Foundation next week.

We believe that the Linux Foundation, they are going to set it up and set up their CICD environment and testing environment by December, mid-December we believe. Again, that's up to the Linux Foundation to make it available. But initially, it will be available for friendly companies who are going to be joining the project. Early next year, in 1Q next year it's going to be available to the public.

[0:36:10.6] SM: Okay. Well, I'll definitely be following along. It sounds like an interesting project, and I can definitely – there's no question that the need is there, but it's again, it's a huge challenge. Network and operations, I think maybe the most famous, maybe the most public proof point, example point here is some of the work that Google has done in this area, where they showed that they were able to reduce their cost of – I don't know remember if it was the overall cost of data center operations, or the heating and cooling costs by some 40%. This is going back a year and change ago. Certainly, there are huge opportunities there for AT&T. How are you attacking that space?

[0:37:00.4] MG: Yeah. It's a key space for us. Number one, we probably have more data centers than many companies that I know today in the United States. We own different types of data centers from small to large to support our customers.

When we start thinking about how does this all come together, first is that all the capabilities in these data centers first have to be software. They have to be software running on cloud commodity hardware. If you don't do that, then it's pretty much difficult to do anything else including artificial intelligence and Acumos. You can't, unless you move to a software world. We've been doing that for the past 5 years and that's part of this ONAP effort that I just talked to you about. We are basically going in that directions.

But where we're applying the technology in our data centers, and we will continue to doubling down in the space, from a machine learning perspective and an AI perspective is not just about energy consumption, which is the example that you gave. Although that by itself is a big opportunity. I don't want to minimize that. But there tremendous inefficiencies one can apply machine learning and AI to really do something cheaper, faster and better. Combine that with human expertise, you could really combine the best of both worlds to doing something that we just can't even do today.

We're not just applying it to looking at energy saving, but there are tremendous opportunities in terms of operating these data centers. There are many functions in these data centers. Just think of we as a network company, we support the world here. There are switches, there are routers, there are firewalls. This is the core core of our network of how packets go from A to B. That's the basics here.

All of these have to be automated and it has to be done in such a way that is intelligent. I'll give you a simple example. At every second of the day in the United States, traffic changes. In some cases, we know and we predict that, in some cases we don't know. Suddenly, there is a big event happening in a particular location, a particular city, in a particular corner that just doesn't happen any day, or any year.

Part of where we're applying machine learning and AI is to really dynamically configure our network and our resources around the clock. I use this metaphor as that, think of like we're building a virtual city that's the United States, where we have limited capacity, but the roads – the roads can become narrower or wider, as long as the physical space is the same. How wide or narrow you make that is can change every second of a day, and that's why we're using machine learning and AI to say, "Well, how much capacity can I put in that second of the day there, because we are predicting some traffic is going to be there over the next one hour or five hours? Where do I take capacity from?" Imagine doing that and optimizing that around the clock. That's a big, big undertaking. It's just one other example where we're applying these capabilities.

[0:40:44.2] SM: Are there some other specific examples or use cases that come to mind within the context of network and ops?

[0:40:51.3] MG: Yeah. Another example is that – so imagine a firewall goes down, what do you do? Well, in the old days, it's failed. Somebody has to literally go there, has to remove it, put a new device in there, new firewall and plug everything and test it, makes it work. Could take days, could take weeks, depending if there are extra hardware available, etc.

Today, that's completely different. It's number one, I can spin up the firewall or that switch at any point of the day in any location, that's number one. That's the vision of what we're going after and we started already executing on that vision. We're reaching about 55% of our target network in terms of virtualization by the end of this year.

Now once I can put these capabilities there, the problem happens is that, "Okay, well what happens now if it goes down?" Well, what we're implementing today is what we call it an AI closed loop, which means that we have the capability of extracting data from this, let's say firewall. We're able to process this data using machine learning. We're able to predict if there are some failures maybe happening or some degradation. We are able to apply policies to say,

"Okay, well we know there's something going on that's bad. We know the driver is this. What do I do?"

It could be that, well reboot the firewall. Then we need an action to go and take that action. Today, that's a manual effort. Tickets are being – we developed tickets, thousands of tickets around the clock hoping that a human, an expert is going to look at those tickets and take an action. But really, a good portion of those tickets are things we already know. We've learned from the data. We know what's a good action, what a bad action.

That's where AI comes in is learning the good action from the bad action, and being able to automate and do this closed loop in a complete automated way, which could be reboot. It could be just move the traffic, there's something chronic about the device, move the traffic and stand-up traffic extendedly, the applications in another device or another firewall. That's sort of, again you could think of the same thing we're doing with security. There are many, many applications like this we are investigating today.

[0:43:16.6] SM: Taking a step back from portfolio perspective, how do you manage the investments across all of these various opportunities that you have?

[0:43:29.0] MG: Yeah. That's a great question. Number one, the reason why we are building a platform – we've done this for our virtualized cloud. Companies are doing that too for their virtualized cloud. We're doing this for our orchestration and automation platform, the ONAP and what we call it e-comm inside AT&T.

We're doing this for machine learning and AI. First, you've got to make sure the foundation is built once, developed by a community in an open transparent way, but it's built once and evolved. It's not changed every time from an investment that's just prohibited. Okay, that's number one. Number two, anything that sits on top of that investment were making sure it is software and making sure that you could take an investment that we already have and drive that cost down with this new world of machine learning and software.

Every opportunity I talked to you about, it's not a vision and it's not something we would like to do. It's something we're doing today, because there is billions of dollars. We have a business case that says, "If we do this through software and through ONAP and through machine learning, we can save 50% of the cost."

We're going after those where there are billions attached and where the technology and the business case shows we can really shave a lot out of these investments, whether it's capital investments or expense.

[0:45:02.9] SM: Certainly, AT&T has more resources available to it than many companies, but those resources nonetheless are not unlimited. What are the biggest challenges that you're

finding in attacking your wish list? I'm imagining that you've got way more places that you'd like to apply this technology than you can actually take on at any given time?

Beyond the platform is again, like we've talked about that and I totally agree that in this situation that you've described, building out a platform that lowers the barrier to entry for other business, or other parts of the organization to build on top of is a great place to start. But then what are the other barriers that if you can figure out how to fix those, you would be able to accelerate your adoption?

[0:45:58.7] MG: Number one is that every company, and we are no different, has limited resources. That's number one. Number two, we're not doing it alone. We cannot do it alone, independent of even if you have the resources, it's you need the skillset and subject matter expertise in different areas, network expertise and functions that we may have some, but not others.

The challenge here is ensuring the ecosystem we work with, the suppliers, the vendors, the partners that we do active engagement with today, all of us are working with the same framework, with the same DNA, with the same platform. That is really how it's a win-win for everybody. That's number one challenge is we've done a lot of progress in that, which led us to where we are today.

That is definitely a challenge is getting everybody to go into one path using something common. Imagine that if every car manufacturers build their car without any specifications, without any blueprint, it's going to be very expensive. But today, all cars come with literally sort of four wheels maybe, they have a steer and they have a seat. There are some expectations that every car is to have, but every car manufacturers compete differently.

Well, imagine if that was not the case, if there were no specifications, then every car has to be built so differently from scratch. It cost a lot of money to build that car. Today, that's not the case and car manufacturers have managed to get all their suppliers and vendors to align. We're trying to do the same thing here. This is a journey. AT&T is a leader in this journey from the networking side, when you look at all the operators.

We went to open source also to bring in other operators into this game and work with them together, so that it's a win-win for both of us. Well, that's a key challenge is harmonization, which is lead us to the relationship we have with the Linux Foundation as a partner to help us drive harmonization worldwide.

[0:48:12.4] SM: Well, you certainly shared a lot off jewels with us. Are there any other thoughts that you'd want to share with the audience on if they're trying to figure this all out, maybe where

they should start or what they should do, or any other thoughts on – I guess, any thoughts that you'd like to close us out with.

[0:48:32.3] MG: Sure. Number one, I think that no matter which company you're from and no matter what investment you have, don't build your own thing. You have to decide what is a value at your company and a competitive edge for you, versus something that is really should be just mainstream. Number one is that be part of an open source movement. Join us in the networking side. Join us in ONAP.

If you are in the machine learning AI side, don't invest in your own siloes. Join us in a community, so that your investment is multiplied by a factor of a 100 with other people's investment. Don't do it alone. That's number one is be part of the open source and really keep your secret sauce, but try to use your investment plus others to build something that's phenomenal.

Number two is as we build a foundation altogether, I think sky is the limit in terms of the applications. We're just touching the surface of that. I know everyone talks about things like voice recognition and facial detection, these are wonderful, these are great applications. Research in this application started three, four decades ago. If you go back to the 70s, 80s, that's how it started. Today, we are at a point where we can commercialize.

There are a lot of opportunities of applications we haven't even thought of. Applications especially as we get to 5G and the internet of things and edge cloud and applications that none of us have even conceived they existed. We would love to expedite that innovations and that inventions, and we would love everyone to really be part of the movement and not to take a sideline and watch. This is a big movement. This is going to change the future of the industry worldwide. Get involved.

[0:50:37.2] SM: That's great. That's great. Well, Mazin thank you so much for taking the time to chat with us this afternoon.

[0:50:42.9] MG: Thank you and it's my pleasure. Really appreciate the invitation. Look forward to talking to you in the future.

[END OF INTERVIEW]

[0:50:52.0] 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 Mazin or any of the topics covered in this episode, head on over to twimlai.com/talk/78. To follow along with this AI Summit series, visit twimlai.com/aisummit.

Of course, you're encouraged to send along your feedback or questions to us by leaving a comment right on the show notes page or via Twitter to @twimlai or @samcharrington.

Thanks once again to IBM Power for their support of this series. For more about the IBM Power Systems Platform for Enterprise AI, visit twimlai.com/ibmpower.

Thanks once again for listening and catch you next time.

[END]