aiworld Conference & Expo

TRANSCRIPT FROM THE PRESENTATION:

The Future of Practical AI

Aaron Edell
Senior Director, Business Development, Ecosystem, Veritone Inc.

Tyler Shulze
SVP, Veritone Partner

TYLER SHULZE: Today, we're going to talk to you about the future of practical AI. We want a level-set a little bit, I know you've heard a lot about the recent trends in machine learning, artificial intelligence. At Veritone, we are really trying to solve practical use cases through our AI-ware platform and we want to give you some insights today about the way that we're approaching the market and hopefully give you some ways to think about ways that you can apply machine learning and I to your business as well.

We'll kick off with just a little bit of a level-set. You guys probably are aware of this but, the world is rich with data, right? We are producing tons and tons of data, more so than any one of us or the entire humankind can basically process. Only a half a percent of data is actually used in a meaningful way. We're intelligence poor as a result of that and really, what we're trying to do is find creative ways to use machine learning, algorithms, artificial intelligence to make better use of the data that we have and to derive actionable intelligence from that data.

AARON EDELL: As you noted on the first title page, we put, "Machine learning is dead. Long-live machine learning," and the point of this is because we need to think about machine learning in a different way and AI is another term for machine learning. Machine learning is an applicable version of AI where you teach it things, it learns. It's a machine and it learns and it applies that learning to use cases in other solutions. When this first re-emerged, 2013, '14, '15, we really saw this big explosion of machine learning as a service, or MLAS. If you haven't heard of MLAS or machine learning as a service, it's essentially when you have an API endpoint where you send some photos and you get back some meta data. This was really useful because, for us, when this first came out, we were really trying to unlock meta data. We had customers with hundreds of thousands of hours of media content, images, videos. All of what was in these videos was locked up. The only way you could get that out of there was to have a human sit there and watch it and literally tag everything that happens. This is a tree. This is Joe Montana throwing a touchdown, I don't know. I don't watch sports. Whatever he does, throws the ball to the other side. That's human work that's expensive and time consuming so it unlocked these capabilities and it automated it, so it really represented a good opportunity for a strong return on the investment.

As such, we saw big growth in these kinds of companies. People were starting to file a lot more patents, it was opening up to a lot more people. Pilots were doubling every year for projects in the business sector and probably more interestingly, people were spending a lot of money on this. This was an opportunity to really change the game.

Right after this, I started to see a lot more interest in the tools that were used to build these in the first place. Personally, a machine box, we used some research coming out of Facebook and Netflix to build some of our models. We use some or all of these tools to build the capabilities that we have, face recognition, image recognition. The use case was really important. These tools were not so important, but that's because they're very, very hard to use, but as they get adopted more and more by developers, we really started to see this economy of sharing the work that was being done. These are all open-sourced. Netflix is releasing research, Facebook. Normally, you'd think they'd want to keep this close to the chest. They were open-sourcing it. They were letting other people build upon it. It's really a very interesting community that we're getting involved with.

TYLER SHULZE: By way of example, what we've seen in just a very short period of time is really a fundamental shift and an evolution in the way that commercially-available APIs from machine learning have evolved. When this first started, the big software companies they were sort of in a race to create these huge general models and they would take hundreds of thousands of celebrities as an example here, right? Microsoft and Amazon and Google and others were creating a model that could find any celebrity that you wanted. The only issue with that is there's not that many practical use cases to find celebrities. We happen to be in the media and entertainment business and so we have a few, but for most folks who are trying to apply machine learning and AI to their business, a celebrity finder is not that useful. The other side of it is these are huge, general models and this was replicated in class after class, category after category. They were trying to create one-size-fits-all models that could be commercially viable across many, many industries. It was a great starting point, but again, we've evolved significantly since then.

I find this slide really fascinating and this is really our experience at Vertitone by way of transcription or speech-detect transcription. When we started back in 2014, 2015, the cost of transcription on a per hour of media process basis was about three dollars and twenty-four cents and it took about three times as long in terms of the media run time to processing time. Processing was three times as long as the media and the accuracy was not that great, honestly, wasn't anywhere close to human levels, but for certain use cases, it was still somewhat relevant and so we started the path of improvement. Over the course of time here in only a few short years, we've really seen a couple of things, right? We've moved now to a real-time architecture and most media and other data streams can be process in real time. The cost of that processing has come down quite a bit, we say 75 cents here. It's probably lower than that on average, now, and the accuracy has shot up. What we're seeing now, though, is we're going to run into what we call a cognitive barrier. You can only get to 100% accuracy. You can only get to real-time processing and what happens then? The other line, which is the cost per hour, is going to get forced down from competition.

In essence, a lot of these algorithms and the processing is becoming commoditized and that's really the core technology itself being commoditized, but in machine learning 2.0, which Aaron will talk about in a minute, there are some new and interesting ways to really make this more relevant and it doesn't mean that the value of the machine learning is becoming less, it means it's becoming more specific to the clients and the use cases.

AARON EDELL: Machine learning 2.0 is really about, "When does this become useful for me? When do I start making money on this?" The point being that that accuracy that you hear, 99 percent, 70 percent, 100 percent, any time you hear that number coming out of

Google or anybody for that matter, what they're talking about is that model against a validation set. That validation set came from the same set of data they use to train. What they did is they took their whole training set so it's a million faces, it's a hundred pictures of dogs, whatever it is they took 20 percent of that and put it aside and they trained with the 80 percent. When they were done training, they ran it against that remaining 20 percent and they get an accuracy. How many times is it wrong? How many times is it right? That number can be very misleading and speech-detect is a great example because speech-detect works really well for certain kinds of voices and certain kinds of accents.

How many people here have Siri or Alexa or something like that, right? If not all of you are raising your hand, you're lying. We've all used it and we've all been frustrated with it when it doesn't recognize us, right? Amazon could tell us their accuracy is 99 percent and we would say, "Yeah, except for the eight hundred times where it played some weird music, didn't understand what I'm saying." That's because that accuracy number came from their validation set, so what does that mean? Well, why don't we train our models on the same data that we want it to run on? There's a great rule of thumb out there. If you look at your training data and you look at an example from your testing data, a human should not be able to tell the difference. If you can tell the difference, then you're probably not on the right path so, really, the implication means not so much about how good the algorithms are and how specialized the algorithms are and what the math is and how many numbers can I put in there and how good it is. Really, it's about, "How can I get my training data into this ecosystem, into this loop," right? This is the key and the reason being is because we tend to get really confused -- I shouldn't say really confused -- we tend to get a little bit confused about what machine learning can do and what it cannot do.

I have been selling and working with machine learning and implementing machine learning for maybe five years now and I've had some pretty doozies of use cases come up. People wanting some pretty interesting-sounding things, but what happens is you say something like, "OK, we can do emotion recognition," so immediately, they go, "Oh, that's very interesting. Why don't I, then, just show all the video of all the people in my city from all our security cameras and then tell you how happy my city is at that given moment?" You start to solve the problem in your head and you don't come back and think, "Wait a minute, can machine learning really do this? Is this really the right use case?" This is one of my favorite examples. The hog is a feature extractor, it's one of those open-sourced technical tools that anybody can use. When you visualize it, if you look at it, you can actually see it actually looks a little bit like a car in this picture. This is what hog thinks a car looks like so when you then show it some ripple in a pond, it's just doing pattern matching. It's not a human, it doesn't have the context. A human can look at this and say a duck, but machine learning might not so you have to think about, "How is my experience going to encompass that when it occasionally gets it wrong?"

The right answer is, again, use training data from your use cases so if I am trying to recognize celebrities in episodes of friends, what I don't want to do is use a picture like this with Jennifer Aniston on a red carpet with different lighting and different makeup and different hair and in different context and then try and recognize her

"There's a great rule of thumb out there. If you look at your training data and you look at an example from your testing data, a human should not be able to tell the difference.."

in this episode. I should train on a screen grab from that episode. I should train on the customer service calls that I'm receiving. I should train on the actual logo in context. This is where that becomes really important because the pattern wants to understand the context as much as it can and the only way you're going to get that is if you train it on that same context.

TYLER SHULZE: One of the things that we found is that there's an inverse correlation between the breadth of training and the accuracy of a given model. If I took any one of you and I had a thousand images of your face and I trained a model specifically on that training data, it would be really good at finding your face 99.9 percent accuracy, really terrible at finding everybody else's face, right? On its own, it's not that useful, but when you start to think about orchestration of these models which is now part of what we're doing at Veritone, there is the possibility and the probability of expanding both the breadth and the accuracy by stacking those individuals models together and using the right models at the right time in the right combination to get the benefit of both breadth and accuracy. We call that process orchestration. We actually have a trademarked software service called Conductor that does just that. This is really the evolution and where the industry is going. We're going to be looking at very specifically-trained models; using these open-sourced tools, readily accessible, trainable by you in your hands in a simple and easy UI; putting democratization around AI as opposed to data scientists working in the lab doing R&D for R&D's sake. This is the transition from R&D to ROI.

Let me give you a little bit more color on this. Part of the process that we envision and have been working on in our R&D lab is really around empowering humans to make corrections and to add a human-in-the-loop element to the entire process and the entire workflow such that the accuracy can be informed from your specific client data running through the machine and the machine will understand where it's doing a good job and where it's not doing such a good job and identify those areas that need to be improved on. We can efficiently now route some of those pieces of media or data back for human review through a crowd-sourcing service or mechanical turk or even have clients themselves that want to do this. They can reintroduce that training data into the trainable models that are on our platform such that the water line continues to increase over time. What we've done now is taken a step away from generalized API-based models and really put the specialization, the training in the hands of every user, but, again, no data science superpowers required. This is part of the democratization process that we're trying to think through and it's actually resonating quite well with our customers.

A lot has been made of the potential of these machines to replace humans and what we like to think about is that this is a human productivity multiplier and not a human brain replacement. I want to give you some examples of real uses cases that we have experienced where we are really helping people to do their jobs more efficiently and effectively, not necessarily just taking them out of the picture entirely. One example is just index large volumes of video. We're in the media and entertainment space, we have a lot of very rich archives. The ability for an editor who's trying to find the shot of Pete Rose stealing second base in 1973 against the Braves on a rainy Tuesday, literally you can now search for if the content is indexed correctly. It helps them instead of going through those hours of tape and trying to find the tape that has the right label on it, they can just pull that data from the system, find the clip that they want, insert it into that retrospective sports piece, and they're done.

Similar use case, transcribe and translate millions of recorded phone calls. There are many, many examples of compliance calls or looking for insider trading or monitoring mortgage brokers to make sure that they're saying the right things. Previously, the only way to do that monitoring was with selective sampling. Literally, one out of one hundred calls might have been listened to and very briefly and probably on double speed so the efficacy of that solution isn't very great. When you can run everything through the machine and you get a searchable transcript, it makes the compliance officers work much easier and now you can assimilate that with texts and with emails and everything else through text analytics. Again, compliance becomes a much easier task.

Similarly, analyzing satellite imagery. We've had some interesting use cases come about where the department of natural resources wants to track the migration patterns of herds of wild animals across large swaths of land. Literally, humans had to sit and look at these photos taken from an airplane and try to identify the tiny specs on those photos to tell whether they were a caribou, they were an antelope, they were whatever. It took hours and hours of painstaking work. Why do that when you can train a model to identify the slight differences in variations in color, in shape, and other things to make their lives much easier.

Lastly, security use case. Rather than having to try to keep track of ex-employees who may or may not show up on your campus, now it's relatively straightforward to do facial matching and facial recognition by taking the photo from their lanyard, putting it into a model, and saying if this person ever shows up again on this campus you might want to alert security because they're probably not here for the right reasons. Again, all of these use cases are really about making human productivity in the workplace more effective and we are surfacing them up through really easy and simple-to-use tools.

This is another interesting use case and something we're quite proud of. In the law enforcement space, we had a use case request to quickly and easily identify suspects. Again, the human piece of this in the past had been literally taking a three-ring binder full of mug shots, paging through to try to find the match to a surveillance camera video in the hope that you might find that needle in the haystack. Now, with the tools that we have available, it's really easy to create a known-offender database or a database of mugshots, use some relatively straightforward facial recognition tools all wrapped into a nice UI for police officers who never really had any experience with machine learning before, but provide them with access to a tool set that makes it really simple for them to find the people they're looking for. Interesting in this case, in only I think the first 30 days of testing this in pilot, we helped a local police agency identify 20 suspects that they otherwise wouldn't have found and that led to several arrests. Those are the kinds of results and use cases that get people excited about practical AI. You can put it in the hands of someone who previously didn't have a lot of technical experience and all of a sudden, it helps them to do their job better. We found that sort of thing repeated over and over across industry segments.

Again, the way that we've approached this at Veritone is to really think about AI as a platform, the application layer as being as important if not more important than the data science layer of the algorithms under the hood. We like to call it the operating system for AI, it's technically a non-operating system, but there's a lot of corollaries to that and one of the things that we aspire to do is leverage the world investment in R&D and try to take the best-of-breed cognition and build it into our platform. Part of the job that Aaron and I do on a day-to-day basis is to try to pair demand from clients and use cases with the available artificial intelligence or machine learning algorithms that are out there. We track about 10 thousand companies around the globe that make the types of things that are interesting to us and every day, we reach out to one or two of them saying, "Hey, we've got a client that's trying to do this and we know you make something that's similar to that so please join us, join our platform, help us surface this and solve this use case."

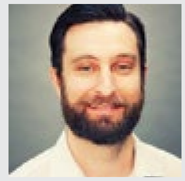
AARON EDELL: I just want to summarize the key lessons that we really want you to take home with you about what machine learning is and why we're here talking about this. Train on data in your real-life use case, OK? Take screen grabs from the movies you're trying to recognize celebrities in. Use your body cam footage, your security footage. Use your data. Make a lot of narrowly-trained models. A model that's trained on just the employees at your facility is going to perform a lot better than a model that's trained on everybody in the world. Humans don't know a million people. We don't have that capability so it's unrealistic, I think, to expect a machine to be able to do the same. Train it on your dataset.

Orchestrate your models in a way that keeps humans in the loop so when something isn't right, teach it the right thing and have that process be embedded in the learning so that you're always learning and that you're preventing model drift and that you're always including samples from your data. This can be accomplished with simple user experience. This is not about which algorithm is the best. This does not require data science to figure out. This is just standard user experience for any kind of application or methodology for using these capabilities. That's what it's about today. The answer to these questions is no longer, "Who has the best data scientist and who can do the best math." It's, "Who's just really good at making apps? Who knows what people think? Who knows how to interact with people and how to get this experience to where it needs to be?" That's really what we want you to take home from this presentation. Hopefully, we made it pretty clear. If you have any questions, come find us at the booth just over there, 107. We'd be happy to go into any of this any further. Thank you.

TYLER SHULZE: Thank you.

"The answer to these questions is no longer, "Who has the best data scientist and who can do the best math." It's, "Who's just really good at making apps? Who knows what people think?

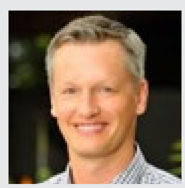
About the Speakers



Aaron Edell

Senior Director, Business Development, Ecosystem, Veritone Inc.

Edell is a veteran speaker and writer on the topics of machine learning, metadata, and content management. He was the co-founder and CEO of Machine Box, Inc., an award-winning startup that builds production-ready machine learning models that anyone can integrate, deploy and scale which was acquired by Veritone in September of 2018. Previously, he helped found and grow Graymeta, Inc., a machine learning and metadata company. Prior to that Edell was at Oracle, Front Porch Digital, Neulion, and SAMMA Systems in various roles from senior solutions architect to product manager. Aaron has published papers on metadata and machine learning and consulted major media and entertainment companies on content management since 2005.



Tyler Schulze

SVP, Veritone Partner

Tyler has over 20 years of diverse professional experience across multiple industries including software, media, entertainment, sports, and automotive. He currently oversees the Veritone partner ecosystem and associated Developer application, forging and managing strategic relationships with cognitive engine developers, application developers, and data product providers on a global basis. Prior to joining Veritone he held a variety of executive and general management roles at NHRA, Motor Trend Media, Fox Sports, Primedia, and eToys. He began his career in Operations and IT consulting at Deloitte. Tyler holds an MBA from UCLA Anderson and a BA in Economics & Management from Albion College.