ON THE PLATFORM

EPISODE: Mighty AI: Data for 20/20+ vision

Host: Matthew Quinlan, Global Geospatial Platforms Lead, Accenture
Guest: Daryn Nakhuda, CEO, Mighty AI

Matthew: Hello, and welcome to *On the Platform*, where we're talking to the most influential and innovative thinkers in platform technology on the hottest topics and trends. My name is Matthew Quinlan. I'm the Global Geospatial Platforms Lead at Accenture, and *On the Platform* today we're going to talk machine learning, and in particular training computer vision algorithms.

I'm joined by Daryn Nakhuda. He's the CEO of Mighty AI, which provides training data as a service to organizations working on developing autonomous vehicles and other applications. For full disclosure, Accenture is an investor in Mighty AI. Daryn, welcome to the program.

Daryn: Thanks, Matthew.

Matthew: So I wanted to start by looking at what are some of the common computer vision scenarios that need training data.

Daryn: Really, you can think about any scenario where you can take human vision and give it a superpower, give it either infinite memory, you know, much broader depth, removal of fatigue. So we see it, obviously, in autonomous vehicles. Self-driving cars is probably the hottest area as far as computer vision goes right now. But certainly there's plenty of other use cases in any situation where there's a camera, either a moving camera or a stationary camera, where things are moving around it and you need to track objects, track what's happening in the scene.

Matthew: Right, so we see things like drone technology, urban air taxis, Amazon Go and cashier-less retail.

Daryn: Lots of great retail use cases, aerial footage. You can talk about surveillance or different applications in which you are tracking a scene over time, and those are all great use cases where right now they may be deployed, but they're deployed with humans. You can make it more efficient as well as higher performing.

Matthew: Right. And we're mostly talking about camera data. And how does that, how does the original data get captured, for example, in autonomous vehicles?

Daryn: Sure. So most of our customers are either manufacturing cameras or an entire solution around a certain functionality, and so they will come to us with raw footage based on their sensors, their vehicles, and we will take that and we will label that according to a specification. So depending on the use case—if it's pedestrian detection, for example, the application would be finding people within the scene, making sure they're labeled, so that might be drawing an outline around each person within a frame or putting boxes around them, and then adding additional metadata that might be helpful to the model as far as what direction is a person moving, are there any other attributes that might be interesting that we want to be tracking over time.

Matthew: Right. So I guess in the case of autonomous vehicles we want perception so we can identify what things are, but also to help decision-making, understand what are they going to do next, so some kind of prediction.

Daryn: That's right. Primarily, when you talk about computer vision data, you're starting at the perception level, but there's certainly several levels below that.

Matthew: And what happens to the data? So we've got this raw camera data and we need to get it annotated. How does that happen in practice?

Daryn: Sure. So annotation is a big word, but it's pretty simple. It's labeling the components that are within an image. So in the previous example it's actually putting that box with here's the coordinates within this image of where the pedestrian is, and for that pedestrian, here's where they're moving over time.

So we might see that same frame, or multiple frames or video and that same pedestrian, Pedestrian A, we'll call him Matthew, moving from, you know, second one to second 10 in and out of the picture, potentially crossing a road. So that's the type of information that'll be used. Then from there it's basically reduced into a form that can be loaded into a deep learning framework where we can take all these attributes and this data and build a machine learning system around that.

Matthew: Right. And as we at Accenture work with automotive and technology companies, we see them working in different locations, so California and other parts of the U.S., training grounds for like AV, but there's also China, Singapore, elsewhere. Do these need different data sets and different training to come up with what the machine learning algorithms need?

Daryn: Yeah, absolutely. And it's really about what is the application and the use case. If you are building a solution that is meant for just one of those environments, data from that one environment may be enough. But really, when you're talking about a general purpose self-driving car that can go most anywhere, if not anywhere, you need that diversity of data.

So that's one thing we spend a lot of time with our customers on, is understanding what data do they have and what data really needs to be labeled. Because especially as we get a broader deployment, you're collecting terabytes and terabytes of data per day, but if you're repeating the same data collection every day it becomes less useful over time. You might catch new anomalies or new pieces of information you didn't before, but at some point you're going to want to extend beyond a one mile stretch of road to really get more examples.

Worldwide, one of the big challenges is road markings, other vehicles, the road signs look very different. As you probably seen when you go to Europe or Asia, there's a lot of things that wouldn't be familiar to an American driver who's only been in America. And you can think of vision systems the same way. If all you've ever seen is a highway sign in the middle of the Midwest, it's going to be very different than being around downtown Tokyo.

Matthew: Right. And how about the kind of rarer events? So even in a geographically limited place a car's going to need to be able to know what to do when a school bus stops or when an ambulance goes past or whatever, which you may not capture just driving to capture regular imagery. How do you kind of capture what you need for those kind of cases?

Daryn: Yeah, so there's a couple of strategies. One, as we get more deployments and more cars on the road, you have more opportunity to capture those. Now, it has to be, you have to be in the right place at the right time. So it's not just that these occur, events happen, because they do happen often, but happening around where your car is in the right place to see that is going to be the hard part.

One strategy that's pretty popular as far as research goes is what they call synthetic data. So where you're taking real world footage, but you're applying an overlay of computer generated content. So that might be weather or lighting conditions, or potentially even scenarios and other objects that go into the roadway.

Matthew: Right.

Daryn: So those can be really valuable. Real world scenarios are the best. And certainly as we see deployments getting that data back into the system is really important. So if you do have a disengagement or you have a complacent which you can, a driver can say I just saw the craziest thing.

Matthew: Yep.

Daryn: The circus train, you know, tipped over. That can be something they can then come back to have that footage be used for training.

Matthew: Right. We do some work with simulation and CGI. What role does that play in maybe taking a core data set and then making it nighttime, or adding snow, or whatever? Is that a valid…?

Daryn: Yeah, absolutely. So that's, when I was referring to synthetic data, that is definitely an area in which there's a lot of research. Certainly the cost of generating variations of your data based on synthesis as opposed to really trying to wade out there and to capture these rare events is a huge one, and the technology just gets better every day.

Matthew: Yeah. So we've got this training data set. Mighty AI and others have created it. What then happens next? How does that get used to train the algorithms?

Daryn: We'll provide that data to our customers, and they will take that, and depending on what kind of deep learning framework they're using, they're basically, they flatten and ingest that data into their training system and build a model.

They'll take other data that is labeled and use that to validate that the model is actually performing better, or performing consistently, so as you train more data you don't end up pulling it away from what was actually higher performing models because you ended up over feeding it, meaning it's gotten too used to just the types of data you have, and not maybe the other kinds of data that you care about it being able to recognize. And after several iterations of that, you will take an evaluation data set, what we call a test set, and then objectively measure how does this do against data it's never seen before.

Matthew: Right, got it. And I guess a lot of the work going on at the moment is in the kind of research or lab, so early production. How does it work once these things, autonomous vehicles, go to scale and production? How does the learning come back into the algorithms then?

Daryn: I think that's still an area that is being discussed, especially within the automotive manufacturers and suppliers. Certainly a hot topic is 5G and better connectivity so that you can have a feedback loop that's easier to do from any car that's on the road.

Matthew: Right.

Daryn: But right now it is pretty much inside of test suites.

Matthew: And does it get to a point where—I guess the more you train the algorithms, can they then become annotators? Does the kind of human annotation get lesser over time, or…?

Daryn: So there's a concept in machine learning called reinforcement learning, and that is certainly like a very interesting spot to think about like how do you not just start recognizing things, but learn processes, learn how to act in certain situations. So what is the right, what is the best outcome, what is the right outcome that you're trying to achieve and how does the computer get to that outcome. I think we're still early days, especially in the perception world, in which, you know, we're just at the first stages of that, but there is definitely an opportunity there.

Matthew: Yeah. And we've talked autonomous vehicles. I guess, you know, the other applications like retail and robotics, do they have very different requirements or is it fairly common?

Daryn: So a lot of the techniques are common as far as how you do object detection or scene segmentation, but the content is very different. And that's actually one of the challenges with open source data sets, is that you really end up wanting to have, for high levels of quality, very specific data sets that tie to your use cases as well as your cameras, your sensors.

So that's definitely one of the things that we find with our customers, is it's hard to find a common piece of data that everybody can use and be happy with and get to the levels of quality they need. But certainly indoor and outdoor conditions are very different. Things like fish eye camera lenses and depth sensors can be different based on their deployment.

Matthew: Right. And how tied is it to the sensor configuration that each customer has? Do they need to have their own imagery data trained?

Daryn: It just gets better. So there's certainly things that can be used generically, and you'll see there's several models out there for things like facial recognition. But when you get down to getting these high levels of precision and quality, training on your own cameras and your own sensor layout.

Well, you know, one of the big things, especially in vehicles, is using multiple sensors, so you might have your LIDAR point cloud sensor, you have your cameras, we have ultrasonic radar. Any one of them by themselves isn't really enough to do what you need to do, but by using them in combination you might get more superhuman power, so not just good eyes, but you also have the ability to see around things, or behind things, or see in the dark. And that's really where that comes in.

Matthew: Yeah, we increasingly see that with clients looking at the whole data supply chain, if you like, and the sensor fusion, and just the challenges of bringing that together, it's hard stuff. As you talk to companies, what are some of the challenges that they tend to face, and what kind of guidance would you give people?

Daryn: One of the things that I think gets under appreciated is just how hard it is to do this kind of labeling at scale. So for a lot of our customers, they are researchers and computer vision researchers who may have developed these algorithms, but they're really reliant on the data. And being able to produce that data to their specification is a challenge.

Maybe you can go and label 100 images yourself, or 100 frames of video, but when you take it to thousands and tens of thousands more, you really have to understand how crisp am I in instructing, what really needs to be captured, and making sure that you, you know, it's the typical garbage in, garbage out problem, in that if you don't exactly specify what you need, you're going to see variance, and that's going to lead to a degradation of your model.

So I think doing things at scale and also having diversity of data and knowing what to label. So one, you know, one of the big challenges is that a lot of companies are gathering terabytes and terabytes of data, and being able to find what is the data that will be meaningful and will really change the performance of this model over time is an art.

Matthew: Well, unfortunately, that's all we've got time for today, but Daryn thank you so much for taking part in today's conversation.

We hope you enjoyed this episode. Please help us get the word out and be sure to subscribe, share, rate and review our series. We'd love to hear from you and hope you tune in again for the next episode of *On the Platform*.

[End of recording.]