EPISODE 57

[INTRODUCTION]

[0:00:10.4] 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. Well, team, I guess I'll just jump straight to the bad news. Last week we shared with excitement the news about a special Halloween event that we were planning for October 30th in New York City.

Well, due to unforeseen events beyond our control the event is now canceled. We were really, really looking forward to it and are incredibly disappointed about its cancellation. If you purchase tickets via either the Eeventbrite or SplashThat pages, you should have been automatically refunded.

The good news though is that you can still connect with us on Monday evening, because will be headed to the NYU Future Labs AI Summit Happy Hour. If you're in New York City we hope you'll join us at the Happy Hour, and more importantly the AI Summit itself. As you may remember we attended the Inaugural Summit back in April and had a great time and delivered some great interviews, all of which we'll link to in the show notes for your listening pleasure.

This year's event features more great speakers including Corinna Cortes, head of research at Google New York; David Venturelli, science operations manager at NASA Ames Auantum AI Lab; and Dennis Mortensen, CEO and founder of startup x.ai. For the event homepage visit aisummit2017.futurelabs.nyc, and for 25% off tickets use the code TWiML25. You can find links to this and more great events on new events page at twimlai.events, and of course this shows notes page at twimlai.com/talk/57.

The episode you're about to hear is the first of a new series of shows on autonomous vehicles. Now, we all know that self-driving cars is one of the hottest topics in machine learning and AI, so of course we had to dig a little deeper into the space. To get us started on this journey, I'm excited to present this interview with Daryn Nakhuda, CEO and cofounder of Mighty AI.

Daryn and I discussed the many challenges of collecting training data for autonomous vehicles along with some thoughts on human powered insights and annotation, semantic segmentation, and a ton more great stuff. You may not realize it, but if you're a long time listener you already know Mighty AI for my interview with their lead data scientist, Angie Hugeback for TWiML Talk number six. It is so hard to believe that that was over 50 shows ago.

Mighty AI was one of the first sponsors of this podcast and it's great to have them back as a sponsor for this series. As you'll hear, the company delivers training and validation data to firms building computer vision models for autonomous vehicles. Their platform combines guaranteed accuracy with scale and expertise. Thanks to their full stack of annotation software, consulting and managed services, proprietary machine learning and global community of prequalified annotators.

We thank Mighty AI for being invited sponsor, so please be sure to visit them at www.mty.ai to learn more and follow them on Twitter @mighty_ai.

Now, on to the show.

[INTERVIEW]

[0:03:47.0] SC: All right everyone, I am on the line with Daryn Nakhuda, the CEO of Mighty AI. Mighty AI is a company that you've heard from on the podcast before. In fact they were formally called Spare 5, and we interviewed one of their lead data scientist; Angie Hugeback, back on TWiML Talk number six just about a year ago. Daryn, welcome to the show.

[0:04:12.4] DN: Thanks, Sam.

[0:04:13.4] SC: It's great to have you on. Why don't we start by having you introduce yourself and talk a little bit about your role at mighty AI?

[0:04:21.8] DN: Sure. My name is Daryn Nakhuda, I'm the CEO and one of the founders of Mighty AI. We've been working at Might AI, as you said, formally known as Spare 5 for about three years and really trying to harness human insights and human power into building better training datasets for artificial intelligence.

[0:04:39.6] SC: Tell us a little bit about your background.

[0:04:42.5] DN: Sure. My background has been in software engineering for about 20 years mainly in internet technology, so everything from e-commerce and communications platforms, through marketplaces.

[0:04:56.4] SC: Okay. Is Might AI your kind of first foray into the AI space, or have you been doing that for a while?

[0:05:05.3] DN: Might Ai is really a first foray in the AI, but really using human insights has been something I've been doing for a while at both startups as well as when I worked at Amazon. I leveraged Mechanical Turk and other platforms to use humans to augment what we could do with our systems.

[0:05:22.6] SC: Okay. Awesome. Since the conversation with Angie, again, just under a year ago, it sounds like you guys have gotten a lot more focused, and in particular you're spending a lot of time in the autonomous vehicle space. Can you tell us a little bit about what you're up to there?

[0:05:42.7] DN: Sure. When we started a few years ago we were really focused on human powered insights for almost anything, and what we realized was what really set us apart was our focus on quality. Like you talked about with Angie a year ago, really building our own models for user reputation and data quality predictions was key to our success. Really, I was resonating with customers who were focusing on building training data or building models where they need a really highly accurate data. That boiled down to natural language and computer vision and really where we saw a lot of focus was on the computer vision side, specifically in autonomous driving, which is a huge field as you've seen. We had a lot of demand there for really specialized, really highly accurate data, so we decided to focus purely on that area.

[0:06:34.5] SC: I've mentioned the conversation with Angie, and you've mention human powered insights a couple of times, and I think I may be taking for granted that folks will have heard that podcast, but I probably shouldn't do that. Why do you take a second to kind of step back and really walk through what you guys do so that we can make sure everyone's on the same page on that.

[0:06:58.4] DN: Sure. We have a platform called Spare 5, which is basically a community of people around the world who we give micro tasks and they are able to perform those. We have a quality control system in which that we can review and manage that both automatically and with other people, and what we delivered to our customers is a high-quality result. They'll come to us with their requirement. For example, a photograph and some requirements as far as what types of things in the photo need to be labeled. In the case of autonomous driving, that might be drawing bounding boxes around pedestrians and vehicles on the road or it could be something like segmenting every pixel of the image into a semantic class. We'll build a workflow and we'll go through the and have humans do that and then using the combination of the humans and our AI, deliver back a result to them that they can use to build their own models.

[0:07:51.9] SC: Okay. You were previously doing this for folks that operated in a variety of market segments, but you've, again, focus more tightly on autonomous vehicles for some time now. Can you talk a little bit about some of what makes that market unique for what you're doing?

[0:08:12.5] DN: Sure. I think autonomous vehicles, especially on the computer vision side is really a great example of what needs to happen in order to build highly accurate models. A lot of the other use cases that we dealt with in the past, there was a lot of flexibility or more subjective insights as to taste, like in retail or something like that where you're much more focused on things that are not life-and-death and safety-related.

With the vehicles, it really is about getting as much data as possible with a lot of diversity as possible and getting it labeled in an accurate way in which we can feel comfortable that we could take this model, train a system, integrate all the sensors in the controls and put a car on the road and have it drive with humans and other cars right next to it.

[0:08:59.3] SC: There are a number of different perspectives on kind of the right way to do autonomous vehicles in terms of, I guess, the different types of sensors. There seems to be one worldview that's kind of very heavily computer vision focus and looks at the cameras kind of the ultimate end-all be-all sensor and there seems to be another point of view that's a little bit more integrative and includes LIDAR and other types of sensors. Do you guys have any perspective on that?

[0:09:32.9] DN: Yeah. Most of the work we're doing is on the image side, camera side, but really our perspective is that in order to have a car drive like a human, it needs to have the senses of a human, right? There's more than just your eyes. That's where other sensors come in, and maybe humans don't have a built-in LIDAR system, but we do have a sense of surrounding, right? It's not just our eyes, but it's sound. When you think about radar or ultrasonic, other contexts that's more 360 that just a front camera or a back camera.

[0:10:06.7] SC: Yeah, I rely pretty heavily on my Spidey sense, which is about as close to LIDAR as I'm going to get.

[0:10:11.5] DN: Sure. There's things that you pick up as you're driving, like you see a person way down the road on the sidewalk and you're going to be thinking about will they cross or not. Maybe that is a camera seeing that, but also the intent of which way are they moving, what are they doing and just what is your experience? Like in downtown Seattle people usually stop at the crosswalk, not really the case the rest of the world.

[0:10:31.8] SC: Right. Can you talk a little bit about where the service that you are providing fits into kind of the broader pipeline that your customers are deploying? Maybe as a prelude to that, you can talk a little bit about the customers that you target and any customers that you can name and kind of what they're working on.

[0:10:56.8] DN: Yeah. We work with a variety of customers, really everybody you could picture in the automotive space. That could be the OEMs, so the car manufactures, the tier 1 suppliers; who are the people who traditionally provided parts but are also now providing integrated systems. Then what we'll call disruptors, so companies like Uber are using autonomous driving maybe not as their core business, but as part of their broader offering. Then startups, who are purely focused on, “We've never been in the automotive space before, maybe we have some individual experience, but now we’re going to go straight after kind of full autonomy.”

That’s a wide range of customers. Most of them are starting out with a car on the road, so equipped with whatever sensors they have. Then they’re taking that data and the requirement is coming from their research team, which might be object detection or it might be semantic segmentation. It could be a combination of them. They give us the raw data, which is video or extracted still frames and the requirements and that we have to develop a workflow in order to give them back the labeled data.

[0:12:03.5] SC: Okay. Object detection that sounds pretty obvious in terms of what that means on face value, but are there nuances that are part of the process there that folks don't generally think of when they hear the phrase object detection?

[0:12:20.0] DN: Oh, absolutely. I think more so in autonomous vehicles than in other spaces where it's not just about what shape is this thing so I can decide whether it's a car or a truck. Even when you see a truck you might have four jeeps, one is a mail delivery van, one is an ice cream delivery truck, one is a passenger vehicle. Those nuances actually become really important when you think about driving patterns. An ice cream truck may have kids running out to it as it drives down the street. A mailman might be driving on the wrong side of the road and stopping very often at mailboxes. Who knows what a passenger vehicle might do.

[0:13:00.0] SC: Right. You also mentioned scene segmentation. Tell us about that.

[0:13:07.1] DN: Yup, that’s right. Sure. Another part of thinking about computer vision is not just what are the objects in front of you, but really what is your context. In this segmentation what we’re doing is labeling really every pixel that's in the field of view, whether that's a road or a marking on the road or a vehicle or pedestrian, different shorts, or vegetation in buildings and curbs. Really, making sure that we have enough information about the different types of things that you're looking at that you can make better decisions.

[0:13:40.1] SC: So every individual pixel gets a label?

[0:13:43.3] DN: That's right.

[0:13:43.6] SC: And the pixels are labeled essentially as objects or is there like a fix vocabulary that you're labeling the pixels with or is it across a broad spectrum of objects?

[0:13:58.4] DN: The taxonomy of labels changes based on our customer requirements, but you can think of them in broader terms kind of as classes or as types of things. It's not necessarily an object, but it's something like the sky or vegetation or an individual car. Usually, from that, what we're doing is labeling everything we can see and then there's additional labeling steps afterwards, so we might breakdown vehicles into, like I was describing earlier, very specific types of vehicles. A lot of these also changes based on the location of the footage, because terminology can change, the types of road markings can change based on what part of the world you're in.

[0:14:39.7] SC: Can you take us a little deeper into the how you do all these? In particular, I'm really interested in hearing how that's evolved from when you were tackling the problem more broadly to doing this specifically for the autonomous vehicle market.

[0:14:57.7] DN: Sure. Probably the best example I can walk you through is on that semantic segmentation that we just spoke about, because it's really hard. Just thinking about that amount of time it would take to figure how to label every pixel in the image regardless of tools.

When we first started out, especially with the Spare 5, we were a mobile only app, and what we've done is we saw that platform but we've also developed a desktop client, or a web-based desktop view. Really, that was about preference. Some people really like working on a tablet with their fingers or with the stylus. Other people really like using the large screen. Ed either way you need to be able to zoom in and really get within a couple of pixel of an edge when you're doing this type of drawing.

Workflow-wise we learned a lot. When we first started doing this, we said, "Okay, we've got a list of 75 classes of things that are in an image, and we built a couple of tools to help you draw polygons so you could go click around a shape and make a close polygons and say, "This belongs to the sky or a car."

What we found was, one, it was really hard to instruct humans on 75 different things at once. If we gave them a long instructional set in quizzed them, it’s a lot to keep in your mind. Two; it takes a long time. Drawing and labeling any one of these images from start to finish might take an hour of your time. Three; if you made mistakes, it was really hard to figure out where the mistake was, how to pick up on it, how to have somebody else come in and fix it or how to have you fix it, so we’re not just throwing out an hour of effort.

We iterated on that process several times as far as an overall workflow. Our first pass about was, well, instead of doing 75 classes at once, we’ll do one. Let's have people focus purely on pedestrians or purely on vehicles, and that helped a lot, because that made them really focus on the instructions as far as where to draw the lines, kind of what counts, like do you go around the tires, do you go around the the bumpers? How tight you have to be? But it’s still is very time-consuming, especially if you think about something like a highway scene in the middle of rush hour where there’s 50 cars within the field of view and some of them really far out on the horizon.

[0:17:13.9] SC: When you switched to that first iterative step, did you go from a model where you would have one worker work on all of the various things in an image to one where the image would kind of pass through steps and be routed to like the pedestrian team and the tree team and the vehicle team, that kind of thing?

[0:17:36.7] DN: Yeah, that's right. I think it’s a little foreshadowing of our next step. Yeah, what we have is from one person one image, to one person doing one class of 75 people roughly for a full image. What we’ve realized was the time between these different tasks was hard to predict and it was still pretty exhausting to do every single car in the scene.

The next iteration was really what we call recursion in our world, but basically we present the image with all the previous activity it's been done to it. If there're 30 cars and 25 of them have already been boxed, we will show you an image with 25 boxed cars and say, “Are there more cars in the picture that haven’t been labeled?” If they say yes, then we give them the drawing tool and say, “Draw the shape around one of the cars. Just do one at a time.”

They will outline a car. They'll label it according to which kind of class it belongs to and they'll hit next. At that point we've taken something that took an hour to do or more to do the entire image to a short a minute or two task to get it right, which allows us both to have that individual unit of work be reviewed both by our automated systems as well as by you our reviewers, and then taking that and aggregating all of those, all the individual cars and all the individual different classes into that final composite image. It just gives us a lot more flexibility into actually how quickly things can run, because things run in parallel as well as the quality, because we have a lot finer grain control as far as what we keep and we don't keep and even how we edit things.

[0:19:23.1] SC: Now, a couple of questions jump out of me. The first is have you thought about making this into like a captcha? It seems like if you can simplify the UI enough it seems like the perfect task to turn into a captcha and just let people who are trying to sign into their bank or whatever do all the work for you?

[0:19:41.6] DN: Yeah. One thing is there are some really specific instructions as far as the different types of classes and labels we want. In some cases, like once something has been drawn and we just need to have you categorize it, that would make sense. Typically there's enough kind of context and we need to train and instruct people on just for them to really do a good job.

[0:20:03.1] SC: Yeah. I realized that part of the value proposition that you are bringing to the table is that you — As opposed to what someone might be able to find with a Mechanical Turk, you've got a pool of workers that you've taught how to do this kind of classification tasks, and so the accuracy is higher and things like that, and so the suggestion is a little bit tongue-in-cheek, but I've been getting a lot of these captchas recently that it will show you a seen and it will say, "Pick all the squares that have street signs in them." It makes me wonder if it's someone like you folks doing basically farming out their object detection to folks that are trying to sign into websites.

[0:20:48.3] DN: Yeah, certainly. I am sure that reCAPTCHA, which is owned by Google, that could be very obviously a use case for them to leverage.

[0:20:57.0] SC: Right.

[0:20:58.7] DN: What we found is, like you said, there is a lot of instruction. We have annotators around the world. We have a really large community. They're in 155 countries around the world. Really, it's us communicating with them and really getting alignment so that when we have somebody doing these tasks the same person is going through the same workflow multiple times and they really aren't just a stranger being shown an image and saying, "Which one has a street sign?" But really given really specific instructions and, also, giving them a way to engage with us, because inevitably as you have this large data set, you'll run into places where the instructions need to be clarified or where they are confused as to, "What do I do if there's a car partially including another car? Which way do I want to draw the boxes?" that kind of thing. That really requires us to interact with them a lot more closely.

[0:21:52.1] SC: Okay. My next question is you have clearly or will have clearly accumulated a ton of labeled data sets here. Is a future step automate — Building some AI models that automate some of these. For example, predict which of the pixels are sky or put a bounding box around the sky and ask the humans to correct as supposed to draw anew?

[0:22:27.4] DN: Yeah, that's absolutely somewhere where we're going to be focusing is how to make the process more efficient. That's both automation upfront as well as assistive tools so that we can make the community just perform better. Right now our drawing tools are very manual where you're clicking every point in a polygon so that you can get a really sharp edge, but there's no reason why with edge detection and other techniques we can make that an easier process for the community members. As well as you said, if we can take that process as described in that workflow of recursion where we have for 40 cars we have 40 people going through and doing each car. If we can skip the first 25, because we can do that in an automated way, that'll be great.

That's not to say we ever want to replace the human, because there's a huge value to kind of having that human eye, that human judgment, because ultimately the eye is only going to be as smart as the people who are training it, but over time if we can kind of up-level them into doing less of the rote task and more of the task, it really involves a human's particular judgments. That's where we want to get to.

[0:23:37.0] SC: Right. Absolutely. I think the safety of the autonomous vehicle will be proportional to the amount of data that has been properly labeled and used in training, and so I think — It just strikes me that there is tons more seen data to be processed. Even if you automated that easy 40% to 80%, there's still going to be plenty of work for the humans to do the more difficult task.

[0:24:13.3] DN: Absolutely. This is an industry where 95% accuracy isn't going to cut it. There're lives in the lines about safety, so you're really trying to get as perfect as you can get, and that's really going to take iteration. That's going to take always having a human in the loop to make sure that there isn't a misjudgment at this point, where we're talking about training and validation before we are even talking about putting this on to the road and into the wild.

[0:24:37.3] SC: Do you think at all about any of the research that's happening around adversarial examples, and there are some that are particularly focused on — I guess it's a bit of a different context from where you're focused, since your focus is on human annotation, but there is some research that looks at things at ways that you can manipulate images so that a neural net will look at a stop sign and see a giraffe or whatever. Is that on your radar at all?

[0:25:05.9] DN: Yeah. It's definitely an area that we've considered both from the adversarial side and in the generative kind of synthetic data side. Really, I think the more that we're tied to what's in the wild, whether it is running into stop signs that have been vandalized in a way that are intentionally trying to confuse models and vision systems, or whether it's about getting a greater diversity.

One of the biggest challenges that auto manufacturers or people who are focusing on autonomous driving have is just how different scenarios are around the world, also rare cases. It's not just about are stop signs being in a different language or a different shape or the road markings is being different. Even the types of vehicles we see on the road, right? Like a pickup truck in the US is pretty different from a pickup truck in parts of Asia.

There're a lot of rare cases. A few months ago here in the northwest, there was like a tractor trailer that turned over in the middle of the highway with a bunch of like sly meals in the back.

[0:26:06.9] SC: Right. I remember that.

[0:26:08.3] DN: It’s like what would you do if you are the car behind — The autonomous vehicle behind the truck as that happened. Currently, even with hundreds of thousands of hours of footage, the odds of getting something like that on tape is going to be difficult or low. Really, I think there is a balance of how do we augment what we have with other scenarios and other things to get that bigger picture of what could possibly happen.

[0:26:36.9] SC: Along those lines, granted that for a lot of the companies in this space, their data is a core element of their IP and ability to differentiate, but are you aware of any movements to create like data consortia, for example, where you OEMs would contribute their data on the agreement that they would get data back so that they may have cars operating in North America and they can contribute their data and get access to data from this based in other geographies? Is that something that — A, is that something that is happening that you're aware of? B, is that something that you might be able to help facilitate?

[0:27:25.9] DN: Absolutely. I think right now there is a lot of secrecy in this industry. Everybody keeps their images, their data, even their requirements as far as what they’re labeling pretty close. They are starting to form more partnerships, companies working together I think both for the reasons you described as well as just — Everybody is working on slightly different angles. If they can leverage each other's to build a solution and come to market sooner or be the first, I think they’re going to embrace that.

Where we can fit in is there's certainly a place in which we can leverage the data that we’ve already labeled and help people distribute that and manage that, so we’re not duplicating as much effort, but we are really thinking about how to build really useful kind of full data set.

[0:28:14.5] SC: Right. Let’s maybe dive back into the process and the lessons learned and how that's expressed itself in technology that you’ve developed. Anything else in terms of specifics? Things that you’ve observed specific to the autonomous vehicle market?

[0:28:36.4] DN: Sure. A lot of things I think could fit a broader market, but really by focusing here it's allowed us to dive deep and not be distracted by what's going on in linguistics and natural language processing versus different parts of robotics and vision. All of these approaches that require humans require a lot of management of the humans.

As far as really working to make sure that we can translate requirements into something that can be understood. Making sure that we understand when people are making mistakes, what is the reason behind it? There's actually a lot of psychology to why do we get that data. Is it because people are being fraudulent? Is that because we didn't explain it right? Is it because we didn't even think about the scenario, or is it because we explained it in a way that they're actually being consistent with what we told them to do, but we are wrong or we misunderstood something.

It's a really iterative process. It's not something where you can just say, "There's a one-size-fits-all tool, drop in your data, use a generic community and get good data out. We've definitely, I think, learned that more than anything over the past few years as far as how much we need to understand really specific requirements as well as how those fit with data that changes over time.

[0:29:51.3] SC: How did your platform expressed those requirements? Are they kind of hardcoded in for each project that you take on or do you have element of the platform that's like a rules engine or something like that? I'm trying to wrap my head around how I might implement something like this.

[0:30:08.9] DN: Sure. It's a combination of many things. As I said, it's been an iterative process over the past few years as far as us developing it. On the instructional side, we spent a lot of time on instructional design as far as just making sure that once we internally have understood all requirements and translated them into something that our community can understand, where both giving them enough information in small enough pieces that they can understand how to use the tool, how to follow instructions for very specific tasks for a particular customer. Even the definitions of how to box or how to draw a shape around, a vehicle might change from project to project.

Making sure that within the context of what they're doing, we're constantly reminding them of the exact rules and testing them. We do have ways to inject known task and make sure they are meeting the right accuracy level as well as getting feedback constantly, so we can tell them, “You're doing a great job at your drawing, but your labels are consistently or sometimes off in some way. You keep categorizing a box van as a pickup, and really they are two different types of things.”

We try to have as much feedback as we can as well as the upfront instructions. The upfront instructions, it’s really — It’s written. It’s showing photographs and showing examples of good and bad. Then even sometimes going in and producing videos that really talk about a nuanced detail that is easier to express with words in motion than it is with just a paragraph and an image.

Part of that too is we have an international community, so making sure that we’re conveying these in the language they understand. There’s no reason why some of these tasks can be done better by one language or one community than another. It's really up to us to make sure that we’re opening it to the right people. If we have a community that speaks — Is natively Spanish-speaking and we give them very nuanced technical instructions in English, it's going to be harder to understand then if we give it to them in Spanish, for example. That's the type of thing that we have to think about whenever we’re doing our targeting as far as who's going have access to this task as well as making sure that between us and our customers that there's alignment.

[0:32:27.0] SC: For a given task, and to be more specific, for a given scene and objects within the scene, how much redundancy is there in the process? Meaning for a given frame of a video, how many times are you asking someone to label a given object before you have that confidence level that it's done correctly? Is there a ton of redundancy in the process or have you managed to kind of filter that out?

[0:32:57.7] DN: There's not a ton of redundancy. Early on in the process we may have most people doing task in order to get a better understanding for the types of differences will see as people do the task. Ultimately, that's part of what makes our system work really well, is that we get more efficient over time and we have less people doing it over time.

[0:33:20.3] SC: Okay.

[0:33:21.0] DN: Unlike a traditional crowdsourcing model where your only quality control mechanism is looking for consensus or asking 10 people and saying, “Six of them agreed, so that must be the right answer.” We try to be a little bit more intentional and intelligent about how we make these decisions using our reputation engines and some of our other internal models.

[0:33:41.3] SC: Okay. This is maybe something that I should've asked earlier, but do you have — Can you share any data points that can help us contextualize the scope of the challenge within the autonomous vehicle space or the volume of data in that space or that you’re working with in particular?

[0:34:02.9] DN: Sure. Right now there's a few cars on the road. I think it’s the easiest way to think about it. Each company has a handful of cars at best collecting data and even one of those cars might be collecting terabyte of video a day, and most that doesn't need to be human-labeled, but there is significant volume especially when you think about the diversity problems we’re talking earlier as far as that's one car on one road, or one set of roads in one area of the world, in the valley or in Germany or in in Michigan.

Really, as these fleets develop, that’s just going to scale exponentially. We’re going to have both the test fleets, which will be hopefully located around the world and collecting different types of data, so not just images, but LIDAR and other sensors. Then when we get into production where we’re going to start looking really for validation and feedback loops, especially when a system gets triggered. If we’re talking about an event-based system, then we have emergency braking triggered. You’re going to want to have a human validate was that the right thing to do or not?

I think over time we end up with more and more use cases that are going to require human insight even beyond just the raw data that's being captured. Part of the art will be figuring out what to annotate or what things need better labeling or what things don't, because obviously we can't take petabytes of data a day and process that in a meaningful way that's going to really improve things.

[0:35:35.3] SC: Right. Do you often get tasks that are incremental in nature? Meaning you've got, as opposed to processing all of the scenes or objects within — Or all of the objects within the scene, the particular use case calls for only the road dividers or signs or things like that. It sounds like that that's a typical thing for you to do.

[0:36:06.1] DN: Sure. Yeah, definitely. Actually, it might be telling about what parts of the problem any one customer is focusing on a given time. Lane markings are obviously a very discrete task as far as looking at exits and dash line, solid lines, rode boundaries, and then pedestrians would be another really good example as far as trying to understand what — In an urban scene, where the pedestrians are located, how they're moving over time. Are the likelihood of somebody to cross the roadway or cross the front of the vehicle or just stand around if it's a bus stop. You kind of have to understand that it's a bus stop where people just stand. They're not going to cross the street. They're not going to move any way. They'll disappear magically in a couple of frames after the bus passes by. There're things like that that I think really are individual areas of focus. Beyond the general kind of computer vision, like building a better eye for the camera is building context and building semantic understanding that I think are involved more of these discrete tasks.

[0:37:14.7] SC: Do you do any labeling? Do you do any — For lack of a better term, I'm thinking of this is like first derivative labeling, like as supposed to saying, "That's a pedestrian." Labeling the pedestrian as walking in direction X or at a speed that you can calculate based on the timestamps on different images and things like that?

[0:37:40.0] DN: Yeah. We definitely do tracking across video, and that's actually — There's two ways to do that. You can either derive it from two still frames or you can play a video, which can sometimes be helpful as far as understanding what else is going on in the frame and just getting all that data at once. Having one person view five seconds might give you more information as far as like if the rate of movement changes, like persons walking and then the cross signal turns to a blinking hand, they start walking faster, or something like that. It's kind of helpful to see that happen, or a car is turning into their lane and so they stop in the middle of the road. It's a little harder to see that when you're talking about an individual frame one at a time even if you're trying to piece out data back together.

There are definitely nuances where it’s not just a box around a person, but, like I said, what kind of person. Do they have a stroller? Are they walking? Are they distracted in some way? And then their orientation, so what direction are they moving? All of that metadata kind of feeds into it where you end up an annotation that's not just an image with a box and coordinates, but it’s an image with a box with the coordinates that has a lot of metadata. It might be related to this point in time, versus the same image in a later point in time that has a lot of shared metadata, but also certain things change.

[0:39:02.5] SC: It sounds like you’re also able to uniquely identify and track not just a person in a box, but person X in a box in frame one across all of the frames in a segment in which they’re visible.

[0:39:19.9] DN: Yeah, that's usually important, to have that instance level kind of tracking. So you can say, “Our car is changing lanes,” or somebody crossing the street or is it just that there's different people throughout the scene. It’s really important to know that, that kind of tracking.

[0:39:35.9] SC: Do you have a sense for who is kind of leading the field in terms of data collection? You’ve mentioned that most of the folks that are doing this have one or two cars out there, but certainly Google has got more cars. At least they've got — They've got a lot of cars that they’ve instrumented for maps that are capturing some of the same types of data. Tesla has got a lot of cars out there with cameras mounted. What’s your sense for who’s got the most data, visual data on vehicles, real-life in the wild vehicles?

[0:40:16.5] DN: I don’t know that I would name a particular company, but really you think about companies that have vehicles in the wild, which might be manufactured in production vehicles or could be fleets. Certainly there are companies that are trying to say distribute dash cams across consumer market so that they can capture video and use it for building autonomous systems while also providing value to the end customer. There’s also things like taxis or Ubers where there's an internet value of that data to the driver, so there's a reason for them to put this device in their car, but there's also the value of the data collection.

I think, ultimately, there's going to be a couple of different strategies. It's not necessarily going to be you have to produce cars and get them out there. You have to produce a way to collect this data that's meaningful for people so that they're willing to do it.

[0:41:10.8] SC: I've not seen the free dash cam if you give us the ability to use the data. Do you know specially someone who's doing that?

[0:41:22.1] DN: There are a couple of companies. I can find the names for you a little bit later.

[0:41:27.4] SC: Okay.

[0:41:28.0] DN: That's one company called Nexar that has a dash cam app. There's another company that's doing it specifically around ridesharing. That's an inside-outside camera. The idea being you're going to have — You're going to see your customers in the back in case there's any situation where you have an abusive customer or an accident. You need to have that liability coverage, as well as you've got your front camera for accidents and that kind of thing.

[0:41:54.2] SC: Formulating that question, I hadn't really thought about all of the dash cams and the various cameras that are mounted on public safety vehicles and utility fleets and things like that. There's just a ton of image data out there.

[0:42:10.1] DN: Yeah. If you think about companies that have been working on mapping for a long time, like you mentioned, the Google Street View and the Google maps cameras, but even beyond that, any fleet or any — All of us who carry smartphones in our pockets and have some apps running in the background with location awareness. That's all validated as far as understanding kind of movement patterns. That alone might not be enough with that in tandem with the camera becomes a hugely valuable, or that in tandem with high-def maps can tell you when there are patterns that are changing.

[0:42:43.7] SC: Is anyone doing anything as far as integrating in visual data collected via drones in this space?

[0:42:51.7] DN: I think that's a whole separate field as far as on the mapping side for sure. On the actual vehicle driving systems, not that I know of.

[0:43:01.2] SC: Is that because you really need to kind of — For the visual data to have the unique perspective of the vehicle to be useful or because it just hasn't happened yet?

[0:43:14.5] DN: I think it's probably a little both. Certainly, companies think that one of their unique advantages is not just the footage they're collecting, but the way they're collection it. Whether it's using multiple sensors in a certain way in certain positions, so where do they locate their cameras. Are they using a stereo camera? Are they using side cameras as well? Wide-field view camera in tandem.

i think there is value to that uniqueness, but also I think we're just going to figure out what the right combination of data is. Obviously, the more we can get, the better. For certain things, like figuring out a lot of big picture view of your current area, it would be great if you had something flying above you the whole time that could see a further distance and a wider range than your eyes or your front cameras might see.

[0:44:05.2] SC: Interesting. Interesting. What questions should I be asking that I might not have asked you yet? Are there other areas that we might want to dig into before we start to wrap things up?

[0:44:17.1] DN: Well, I think we've covered a lot of good topics. I think we talked about how people approach this problem. That might be interesting. Before Might AI, when you talk about other crowdsourcing or you talk about doing it yourself, one of the biggest challenges is about the quality control. Like I said, the instructions and all of that. Even when you're trying to do it in-house with your own team who knows all the requirements and all the instructions, it's really about that scale and diversity.

I think really iterating that or kind of thinking about the fact that this data in Seattle is different from the data in Detroit, which is different than the data in [inaudible 0:44:57.8] or Singapore or any of these parts of the world. I think that's pretty key. Like you asked earlier about how to distribute, how to collect more data, it's not just like drive the same car in the same route over and over and over and over again. I think you do need that for a little bit. Really, it's about the diversity and then the understanding, because your labelers in the US might not even recognize what does it mean when I see a zigzag line on the side of the road in Europe where we go, "Oh, that's a no parking zone," or "That's a merge area."

There's so much context and so much localization that I think is easy to overlook. Even if you're somebody who's traveled and you know that you need to go learn the rules of the road somewhere else, it's like a lot of things that a human can adapt to that if you're thinking about a system that really it's just looking at what it sees and not having the higher level of understanding, it's really a hard problem. I've never seen a zigzag road. Does that mean I have to like slalom my way down the road as a car, or does it just mean like stay away from that line?

[0:46:03.5] SC: Let the car figure it out.

[0:46:03.8] DN: Yeah, it wouldn't be surprised to see some autonomous vehicles see one of those markings and just go crazy as far as what it's supposed to do.

[0:46:12.0] SC: Yeah. It's interesting, there's so much of this problem space that benefits from having intelligence and the human — Or the computer intelligence and the human intelligence kind of melded together. It's not just this training data labeling problem that we've talked about, and you've made a very strong case for the power of combining human insight with automated tools. Even within the vehicle itself, there are folks doing research on how the car can benefit from the input of just looking at the driver and understanding what their state is, what they're looking at, things like that.

[0:46:59.1] DN: Absolutely. Yeah, I think ultimately when we get to a point where we have full autonomy, we're going to be in a safer world where we don't have distraction, or distraction doesn't matter. It's okay to sit there and stare at your phone if you're not the one driving and you don't need to be the one who's ready to grab the wheel.

We're talking about early stages where you need to be attentive and have your hands on the wheel or close to it. As we get further down the road, you take the best of the humans, which is the judgment and the vision and the decision-making processes and you take away the fatigue and distraction and the things that are going on in our lives and make it hard for us to stay focused, and I think you're going to end up in a better place.

[0:47:40.5] SC: Awesome. That is the hope, and the vision behind autonomous vehicles is as much as people talk about, for example, the economic issues associated with deploying a bunch of autonomous vehicles in terms of labor and things like that, the promise to stave off the huge numbers of vehicular-related deaths that occur around the world, that's just huge.

[0:48:09.0] DN: Yeah, absolutely.

[0:48:10.2] SC: Awesome. I really enjoyed this conversation. Thank you so much for taking the time, and I really appreciate getting a chance to catch up with Mighty AI and hear about what you are doing in this space.

[0:48:23.8] DN: Great. Thank you, Sam. Have a great day.

[0:48:26.0] SC: Thanks, Daryn.

[END OF INTERVIEW]

[0:48:30.3] SC: All right, everyone. That our show for today. Thanks so much for listening, and of course for your ongoing feedback and support. For more information on Daryn or any of the other topics covered in this episode, head on over to twimlai.com/talk/57. To keep track of this autonomous vehicle series, visit twimlai.com/avh2017.

Please, please, please remember to send us any comments or questions you may have for us or our guests via twitter @twimlai or at @samcharrington, or leave a comment on the show page.

Thanks again for listening, and catch you next time.

[END]

