EPISODE 58

[0:00:10.6] 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.

We are back with our second show this week, episode 2 of our Autonomous Vehicle Series. This time around, we're joined by Jianxiong Xiao of AutoX, a company building computer vision centric solutions for autonomous vehicles.

Jianxiong, a PhD graduate of MIT CSAIL Lab joins me to discuss the different layers of the autonomous vehicle stack and the models for machine perception currently used in self-driving cars. If you're new to the autonomous vehicle space, I am confident that you'll learn a ton. Even if you know the space in general, you'll get a really interesting glimpse into why Jianxiong thinks AutoX's direct perception approach is superior to end-to-end processing or mediated perception.

Our Autonomous Vehicle Series is sponsored by Mighty AI, and I like to take a moment now to thank them for their support. Mighty AI delivers training and validation data to companies 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 embedded services, proprietary machine learning and a global community of pre-qualified annotators. If you haven't caught my interview with their CEO, Daryn Nakhuda in the last show TWiML Talk #57, please make sure to check it out. Of course, be sure to visit them at www.mty.ai to learn more and follow them at @might_ai on Twitter.

Before we jump in, if you're in New York City next week we hope you'll join us at the NYU Future Labs AI Summit. As you may remember, we attended the inaugural summit back in April; had a great time and shared some great interviews, which will 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; Davide Venturelli, Science Operations Manager at NASA Ames Quantum AI Lab; and Dennis Mortensen, CEO and Founder of startup x.ai. For the event homepage, visit aisummit2017.futurelabs.nyc. For 25% off of all tickets, use code TWiML25. We'll be at the summit happy hour on Monday the 30th and the summit itself on Tuesday the 31st, and we look forward to meeting you there.

Now, on to the show.

[INTERVIEW]

[0:03:01.4] JM: All right, everyone. I am on the line with Jianxiong Xiao. Jianxiong is the Founder and CEO of AutoX, a company that is doing some interesting things in the autonomous vehicle space. Jianxiong, welcome to This Week in Machine Learning and AI.

[0:03:19.0] JX: Thank you, everyone.

[0:03:20.2] JM: Great. Why don't we get started by having you introduce yourself and tell us a little bit about your background and how you got interested in autonomous vehicles and ML and AI more generally?

[0:03:34.5] JX: Yeah, sure. Sounds great. I got my PhD from MIT. I spent four years at MIT in the Computer Science and Artificial Intelligence trying to get PhD. My research area is in Computer Vision and Robotics. I have been working in this space for quite a while. After I graduated from MIT, I went to Princeton University as a professor. I was the founding director of the Computer Vision and Robotics at Princeton University at the Department of Computer Science.

Our research is about trying to make computer see and enable them to interact with the physical world. For example, these days we have image that from camera or from other sensor, this is called sensor input, and then we'll try to pass what's going on and in order to get a understanding of the physical world about the traffic situation, about people walking around and in order to make sense of the world.

After that, we can interact with the physical world, we can design robots. So you know that, we were designing all kinds of robots including of course my big self-driving cars as well as a smaller robot walking around, moving around inside an indoor space and have robot hands to grab objects and so on.

For example, I will participate to Amazon picking chance last year with a robot arm, together with a team of experts from MIT. We have a joint team together. We get a pretty good score in the final competition for robot picking challenge as well.

[0:05:07.5] JM: Wow. Is this the first time you've participated in the picking challenge?

[0:05:11.8] JX: It was the first time, yeah. But also the last time. After that, I started this company AutoX working on self-driving car. We're being very focused. So right now we focus on trying to make the self-driving car technology really good enough for practical usage.

[0:05:34.0] JM: What I've learned about the company just from some of my background reading is that you are really focused on trying to enable self-driving cars based strictly on vision-based technologies, as opposed to LIDAR and some of the other sensors that we think of like the Google self-driving car. Is that the case?

[0:05:56.7] JX: Kind of. It's not exactly like that. But we are – well a solution I would say is camera first solution. We are not against any other sensor. We are very open to use any sensor. But at the same time we primarily focus on using camera as our primary sensor.

One of the reason is that – there are two reasons, one is the cost factor. Cameras these days, they are very low-cost. But other sensors, such as high-resolution light they are very expensive. If we have to make a product that you can already be used by average citizens, by everyone, it has to be low cost so that financially it actually makes sense to have the self-driving car. Because if a self-driving car is way more expensive than a car plus a few Futang driver, then it doesn't really make any sense, practically.

So for us, we are a camera first solution. The other thing is also technically it's not just the cars. Technically, a typical camera has very high resolution, but even a very high-end LIDAR. For example, a very expensive $80,000 US dollar LIDAR, is still [inaudible 0:07:07.8] 64 bit vertically. The resolution of a high-end LIDAR is simply too low so that it cannot be used for many situations, such as that city and city downtown level 5. Urban driving, it has to be able to recognize those subtle details, small objects, complications in order to be able to drive safely. High-end LIDAR simply doesn't already have enough resolution compared to a camera to do this task.

[0:07:39.2] JM: You said that LIDAR with the 16 – did you say 16 by 16 resolution, or how did you –

[0:07:47.1] JX: LIDAR. A lot of LIDAR these days, they are spinning LIDAR, so they have 360 degree coverage horizontally. But vertically, they have a fixed number of beams. The LIDAR was a recording tool, it's the highest in LIDAR is 64, 64.

[0:08:03.1] JM: 64, got it. You said that was – how much did those units cost?

[0:08:08.4] JX: Those unit cost $80,000 US dollar now.

[0:08:11.9] JM: $80,000. Wow.

[0:08:12.7] JX: Yeah, 80. Actually, it's not. The other problem is it's not automatic weight. That means if you keep using the LIDAR in high temperature, in low temperature, it's going to break in a few months. But for any [inaudible 0:08:26.9] can be installed on the vehicle. Actually people is betting it has a lifetime of like 15 years.

[0:08:34.2] JM: Okay. Also on your website, you talk about your goal being to try to enable self-driving cars with just $50 worth of cameras. In fact, you got off the shelf webcams in a picture mounted on the hood of a car. I think they're the same one that I have, the Logitech C920.

[0:08:54.7] JX: Yes, of course. That's not for final production. But yes –

[0:08:57.8] JM: Of course.

[0:08:58.3] JX: For the initial experiment, we're actually using those Logitech webcams. They're very low-cost. But to moving towards production, now we're upgrading our camera with other camera that is not webcam, but that actually a similar price level. They are not much more expensive. Yeah.

[0:09:16.8] JM: Okay. Maybe we can dig into what are some of the things that you're doing, or the ways that you're thinking about the problem that enables you to take this camera-first approach?

[0:09:32.1] JX: Yeah, that's a very good question. One thing that we try to emphasize is camera actually have a lot of potential. For example, a lot of people say that then can your camera base system drive at night? Actually yes, because a lot of camera these days, the feature is actually much better than even human eyes.

What is actually missing for the camera-based solution is really a very good software. It's very sophisticated, very advanced AI algorithm. That's what it's missing to make the camera-based solution reliable.

For example, a human use the guide our two eyes, we don't have a spinning LIDAR on top of our head shooting lasers. But we can still drive very safely. That's what is missing is really about the software. This is where our clean innovation comes in.

Whereas, in our company most our software engineers, our researches, were walking in this space trying to develop a better advanced perception system, as well as a very robust printing system and decision-making system, in order to what to gather side-by-side to have a robust solution. So maybe I think start with the general pipeline of the architecture first.

So there are three major steps to self-driving, in the autonomous driving software steps. One is the perception, the other one is the printing and decision-making, the third one is the control. Possession is referring to the part that will take the image and other sensor as input, trying to catch what is still going on in the physical world, get the traffic situation and try to have the software to understand, "Okay, this is an object. This object is moving and this is a traffic sign and this is a traffic light and get the sense up the wall." That's the perception part.

Then after that, that's decision-making part. It's that given now the computer can understand what's going on in the physical world. Now the computer need to make a decision. Should the car stop or should the car go, how fast it should go, how much it should turn? All these are the decision and printing part.

After it made a print, now finally we need to execute a print. That's the control part. We need to control the vehicle actually going to execute a print and behave accordingly. That's the control part. These are the three major building blocks, three major step in a full step software for autonomous driving.

[0:12:05.9] JM: Okay, perception, planning and control.

[0:12:07.8] JX: Exactly. There are different schools how to do this, there's one approach that – okay, it's my traditional Google-based approach, [inaudible 0:12:17.4] base approach. You try to do every step separately, completely separately.

The perception part we'll focus on making a very good understanding of the physical world, make sure everything is perfect. In particular, you try to get I will say at curricular level, a pixel-wise level, that for each image or each pixel in the image, they try to make a perfect understanding of what the object it is, that is this pixel belong to a car, or is this pixel belong to a room service to have pixel-wise understand it.

Then given this pixel-wise understanding, they try to have a 3D understanding of the physical world, in order to follow to support that decision-making and planning. That's one typical approach, which we call mediated perception.

[0:13:06.6] JM: Mediated perception?

[0:13:07.4] JX: Yes. That's another approach that is recently populated by a media, is trying to end-to-end, everything end-to-end together. Remember we are talking with a three-step, their perception, there is printing, there is control.

The immediate approach is try to fuse everything together. They no longer use a modulized approach. They just put everything together into a huge gigantic new network, so they'll take the sensor, such as the image at the input, and the new network just output and ready to control and how much work you should apply, how much student talk we should apply to the student well. That's the end-to-end approach form and media.

Now both the push have a lot of problem. I would say at least some problem. Maybe we can talk about end-to-end first. The end-to-end first, the problem is everything is working, I would put a box. It's very difficult to provide input. Like for example, if a computer drive a car to a intersection, now you can turn left, now you can also turn right.

But it's actually very difficult to tell the computer to turn left or right, because the computer will look at this intersection and you will make up of is mine, because the software box just – then it will automatically tell you, “Okay, I’m going to turn right. I’m going to just turn it.” It’s very difficult to control what’s going on inside. Then everything is just that box. That’s one major job out the end-to-end approach.

[0:14:35.2] JM: If I can jump in, that’s an issue I’ve never really thought much about is when you’re building systems that are controlled by neural networks, the ability to for example override the system or provide user direction, is that like a software engineering challenge? Like you have systems that take the neural network input and take the user input and just prioritize the user input, or is that like a network – neural network challenge where you’re providing the neural network input in and you have to train the system to prefer it?

[0:15:14.8] JX: I would say it really depends on the detail. For this kind of end-to-end approach, if everything is to end-to-end, completely end-to-end, then that is not just a software engineering problem. It’s already a more research integrated and neural network problem, because the neural network decide everything for you. You don’t really have a choice. What happened to computer, the neural network decide, that’s the result.

But if you are able to use the neural network in a more modulized approach, use the neural network to do certain tasks and the other part to do – another neural network to do certain tasks, and then eventually you have some way to combine together. In that way, the user actually can have more input.

That’s a very good question that points out the one drawback with a completely end-to-end approach, is that now the computer decide where you are going to go, where you are going to stop. That’s that. Obviously, not usable for us. Yeah, and another problem for the end-to-end approach is the amount of data required to have this system up and running. You can imagine that to a coupled space, to train a good neural network, we will need to couple pre-match all the used case, all the potential traffic scenario in the training data in order to train the neural network to behave smartly.

But this kind of approach is very difficult, because you can imagine that even in the same row, in the same row intersection for example, all in the same highway merging point. That could be many different kinds of traffic, right? That could be different numbers of cars. Those car could be a different position. Each car can have different size. Each car could have different color. Each car can have different speed. Each car can have different reaction time. All these are different and now we need to couple the whole space. We need to have enough training data to put forth all the space.

This is still assuming at the same traffic merging point. If you have different row at different row condition, at different end with that different – that makes the number of training data, the requirement is so big, that you probably even like the whole human society capture data for thousands of years, we may not still have enough data to train a good neural network to cover all the space. Yeah, that's another typical drawback of the end-to-end approach is the amount of data it require is really gigantic.

[0:17:36.9] JM: I mean, just as maybe as a counterpoint, I mean the impression I'm getting from folks that are doing things that are more like the end-to-end approach, Invidia and Google is that they're making a lot of progress. I'm not getting the impression that they think it's going to take thousands of years to train these systems to be operable.

[0:18:00.7] JX: That's not really true, because the Google approach they are not end-to-end. No, we are talking –

[0:18:05.2] JM: Oh okay.

[0:18:05.6] JX: The Google approach is the second – the next approach I'm going to describe is the mediated perception approach. They are opposite of end-to-end. They cut the end-to-end into many, many different small step; into so many steps basically.

Each step, they will do something very – just very, very small step. The mediated perception approach typically is that take the sensor input, and remember, the end-to-end is about merging the perception pointing the control all into one single step.

Now the mediated perception is different. Not only they separate them into different step. Even each step, they separate to many sub-steps. Like for example, when Google is approached in this kind of approach, they take the image as the input, and then they try to get the pixel-wise recognition of each pixel in the image. That's not usual for driving. But that's the first step.

But then after that, they convert this pixel-wise segmentation, reach out into a more 3D understanding of their work. For example, they will give a 3D bounding blocks to contain each car. Each vehicle you will have a box to contain the car.

[0:19:20.1] JM: Meaning you've got from your LIDAR sensor, you've got a point cloud and you've got from your image a two-dimensional view that identified that some two-dimensional set of contiguous pixels as a car, they would fuse those to determine a 3D bounding box for the vehicle based on both the point cloud and the image data?

[0:19:48.5] JX: Yes. That's right. But already, you can see about this and there's a lot of information in this process that we're trying so hard to get. They are not particularly useful. For example, the height of the vehicle. We don't really care, right? No matter how tall the vehicle is, we don't want to hit them. There are a lot of information in –

[0:20:15.2] JM: But you want to know how tall – what the clearance is for a bridge that you're trying to cross under, or an overpass.

[0:20:21.5] JX: Yeah exactly. Sure. So that's certainly information that's useful. Certainly information that are not useful for us to try. That's the point exactly I'm making. In Google's approach, it's the opposite of end-to-end. They tried to get everything, no matter if it's useful or it's not useful. They all get it out. You got it whether it's going to be useful or not.

I will say this approach is safer, but it is overkill. There are lots of redundancies that we can squeeze out. Because every bit of information we extract, there's always a cost. There's a cost of computation on both. If you see a lot of information that are not useful, you waste a lot of computation.

Second is also a lot of engineering and research time, because if we spend so many engineering efforts to get those information and actually they are not being used. It is a waste of time as well. Certainly is a waste of churning data, because they didn't even get a lot of churning data with a lot of heavy annotation in order to get something that is not really useful. So that's another problem.

AutoX will come in between. We review that mediated perception have some advantage, but you may be overkill. We see the different problem on end-to-end perception, end-to-end approach that are completely end-to-end, there are a lot of problem but the good thing is the most simple and more elegant.

So we design something called the graph perception, which will fall in between. It's that we are trying to only get those information that are useful for driving. We're not getting those that are usually for driving. A lot of information that are in the mediated perception approach they get out is useful for other tasks that if I'm not – if I'm a bird, I'm flying around, definitely I need to know how high is the car.

There are a lot of useful information for other application, but not for autonomous driving. So for our approach, we try to identity the decent information that is useful for autonomous driving. We ignore those that are useless for autonomous driving and we only focus on spending the computation power, spending the engineering effort, and spending money on gathering training data for those useful information. We call those useful information affordance indicator.

[0:22:40.6] JM: What indicator?

[0:22:42.0] JX: Affordance. For example, I can give you an example. This is terminology allocated for robotic only. For example, if I give you a mug that you can drink water, whether the mug you can have a handle, right? Typical mug have a handle. The handle for you to grab the handle, so that you can raise the mug. Affordance means the environment of the object allow you to, support you to do certain action. Intelligent agents to do certain actions. So, that's what we call affordance.

In the autonomous driving scenario, is the same – is the car's traffic situation can afford you to do certain task. Like for example, is the traffic – now is a traffic jam, that means the affordance of this car in traffic situation cannot afford you to speed up your car and into a 60 miles per hour. Autonomous driving what we actually need is we need to get the affordance. Can the kind of traffic situation, can the kind of low condition, can a car's physical condition allow you, can afford the autonomous driving car to perform certain actions? This is the list of essentials things that way we really need for autonomous driving.

[0:24:00.2] JM: I guess one question that comes to mind for me is that affordance as a key metric seems – sorry, what is it that I say? It strikes me as it's like it's a planning metric. Like if I have a route that I'm trying to pursue and I want to plan my next step, whether it's change lane or turn or something like that, does the current situation allow me to do what I want to do?

There's also a requirement that these vehicles be reactionary to things that happen. I'm wondering, when I think just the way affordance sounds, like it's not necessarily as applicable in those types of scenarios, is that the case?

[0:24:51.4] JX: In the narrow sense, maybe yes. But in the general sense, because when we say affordance, it's not a static thing; it's a dynamic thing. Like for example, in the more computer science language, the computer is making a decision, 30 frames per second. That means every second the computer is making 30 decision. That means the computer is changing their mind quickly, very quickly, plenty faster than a human being.

[0:25:18.9] JM: So if the car cuts me off, then I still need to figure out if I can afford to go straight for example? That's happening at 30 frames per second.

[0:25:29.6] JX: Yeah. Then suddenly, the affordance becomes very reactive, because if the situation changes a little bit, the affordance changes, and then you're basically being very reactive. In the general sense, affordance is referring to all these reactive behavior as well.

[0:25:46.7] JM: Okay. Thanks.

[0:25:48.2] JX: Yeah, forming up our stories what we do is we take the existing autonomous driving approach and we squeeze out and we see which part is already essential, which part is – which affordance is really necessary for autonomous driving. We are focusing our energy on making those very reliable, so that we can have very robust autonomous driving solution. That's like basically the unique power of technology.

[0:26:16.6] JM: It sounds like then in the mediated perception world, they are – are they ultimately trying to get to affordances as well, but they haven't pruned the universes of affordances that they care about, or is there not really the concept of affordance there?

[0:26:40.0] JX: I believe that's a concept of affordance that is just like when they designed this system, it probably – it was like a decade ago, this car system – at that time, they didn't seem too much about this. They spend a lot of time probably focused on other aspects.

The system design was a little bit no longer the greatest way, they're no longer the most elegant way to do the technology. I also believe that maybe in the future when they were also mixing about this, people are smart and they will also improve as well.

[0:27:16.2] JM: When we're talking affordances and this direct perception trying to focus on only the most relevant affordances, is there an enumeration of those? Are there 10, are there 20, are there hundreds? Does that question make sense, or is it more like, is it less a high-level concept and more something that's implemented in the software that's like some vector of affordances that's determined on the fly?

[0:27:48.9] JX: Yeah, that's my sense. It does if you enumerate on this, it's probably less than 200. It's 100 something. Of course, a different traffic situation, it's not like you have the whole 100. Always does, always a subset that made sense.

We'll classify the traffic scenario into different subsets, and then each subset we will be getting a smaller number, a perception indicator, the affordance indicator in order to drive the car and to have the car behave appropriately.

[0:28:22.8] JM: Okay. And so since you called this direct perception, does that mean that the planning and the control layers of the stack remain the same and they're just getting inputs from this different kind of perception, or did those also change to accommodate direct perception and affordances?

[0:28:42.7] JX: Yes and no. Yes in the sense that if we just focus on talking about direct perception versus mediated perception, yes this part, the difference is only on the perception. There's another dimension on technology that's different from other companies, we mentioned that this at the beginning of the conversation is we are focusing more on using camera, instead of LIDAR.

The mediated perception, direct perception, they can both apply to LIDAR and camera. But if we are talking about camera, another way of difficulty because for camera, because it's not an active sensor, it's not shooting out a laser, the distance measurement is usually more noisier. There would be more noise in the distance measurement.

We have to model the uncertainty of the optic distance, optic speed and so on. That makes the decision-making, printing and control part has to be modeled by us as well. Therefore, it's always that decision-making and the printing part we'll also spend a lot of time making to take the uncertainty from the perception needs into account as well, in order to have a very robust autonomous driving system.

[0:29:59.7] JM: When I think of a system that is looking at images, trying to calculate distances from the car to objects in those images, and then applying another layer of uncertainty, that calls to mind some type of bazian type of system. Is that what you're doing underneath?

[0:30:21.0] JX: It's not exactly bazian, but bazian is one of the many way to incur uncertainty. But in-house we are inventing a lot of other ways to incur uncertainty as well. It's not necessarily fully bazian firmware, because technically bazian firmware will have some technical requirement of being bazian. But at least, we're still taking the uncertainty into account.

For a lot of other systems, for traditional autonomous driving systems, they actually don't take uncertainty into account, that printing and decision-making is completely decisive, if this is going to happen, then it will happen. It is a pure if else statement only. But for us, socially is much more robust because we consider the uncertainty of the perception itself.

We take uncertainty into account, with – assuming the perception we found is not perfect and we are able to make use of the uncertainty estimation. In order to make a frame that is reliable, even if that something still happen.

[0:31:28.2] JM: When you talk about the if-else kinds of construct, is that specific to mediated perception? Like end-to-end, they would just be a neural network that's doing something, that doesn't involve if-else, but with a mediated perception, you've got some control system that is taking in perception inputs and planning as based on if-else. I guess, I don't think of these systems as very if and else, as very traditionally rules-based as opposed to based on train networks.

[0:32:07.4] JX: Yes, you are right. When I say if-else statement, it is referring mostly to mediated perception system. The reason why it is that is because I would say 99% of the companies working in this space they are still working within mediated perception. The end-to-end approach is still mostly a research – on the research side. It's still very difficult to be practically used.

[0:32:33.5] JM: Okay. At what layer of the stack do you find the if-else, the rules-based parts of the system in a mediated perception approach?

[0:32:47.1] JX: It's usually on the decision-making, after the perception and before the control, the printing or decision-making. Usually people have to manually write down all the rules at least hundreds, thousands of rules.

[0:33:01.1] JM: For example, if a car is kind of encroaching on my lane from the lane then turn right, like what's the granularity of these rules, I guess is what I'm trying to wrap my head around. Can you give me examples there?

[0:33:18.4] JX: This granularity of the rule is very detailed. It's not just so general. It's more like if the car is approaching you from the left lane, you need to estimate the size of the car, the distance of the car and the speed of the car and the acceleration of the car in order for you to decide what to do accordingly.

It depends on different speed, different size, different load structure, is it right-turning, left turning right, is it on the street, is it on the curb, you know? They are all different. So you have to incur all these information, and they have many different combination in order to incur all these rules. That's why there are so many rules have to be rewritten down.

[0:34:01.3] JM: Right. I'm trying to put together these three approaches that you've mentioned, end-to-end, pixel-wise, direct perception. It strikes me that you can have a system that is – you know, there is still more degrees of flexibility of kind of inserting machine learning into different layers of this.

For example, you know with mediated perception, a company could start with their mediated perception system and swap out the planning decision-making with the train neural net for example, without changing the way the perception works. Where does that fall apart? Why aren't people doing that?

[0:34:51.5] JX: It's possible. That's possible, but just practically very difficult to make it work. Right now the neural network is still performing very well, mostly on the perception step. For the other decision-making unsolved, it's very difficult to still have the neural network working.

There have been some research, for example using the reinforcement only – [inaudible], to do these other tasks. But it's not at the level of maturity that most people are willing to use for production here. It's still mostly in the research stage, it's not ready for the real product.

[0:35:35.2] JM: Even non-deep learning machine learning models. I guess, what I'm trying to wrap my head around is intuitively after you've gone through the perception step, you've got a set of features that represent what you've learned about 3D space around the vehicle, the vehicle dynamics and all that. Is it that the dimensionality of that is too high for either traditional machine learning models or deep learning, or is it something else that is really the main challenge?

[0:36:19.4] JX: The difficulty is not about the dimension or the size of the data. The difficulty is about the space is not a fixed mapping. Let's put it this way, for neural network a lot of traditional machine learning, such as the [inaudible] machine, what they are trying to learn is a function. You give me some input. I have a mapping to man output. So it's a function. You change your input, my output changes, and so on. But when we're talking about robotics, or we talk about autonomous driving, specifically it's very [inaudible]. What is going to happen in the next step depends on what I do now, that if I speed up my car, suddenly the car in front of me may get scared and may speed up as well.

Since I'm not predictable, it's not just me is changing. It's not a fixed mapping. Is it everything going to happen in the future depends on what's happening now. That's a sequential causality in between. That makes the fixed mapping function, as mapping have to [inaudible]. That's not very – not really applicable to this kind of application. That's why people have to invent something newer, more fancier stuff, try to do this.

[0:37:41.5] JM: Okay. So tell me a little bit about the progress you've made as how far along are you?

[0:37:47.4] JX: Yeah, we are a young company. We got started 13 months ago, just a little bit over a year. In the past year, we already made a lot of progress. People from our side, AutoX.ai you may see that our car already saw some initial prototype driving on the street, doing a lot of demos, can drive safely using camera to achieve almost all the driving behavior that we require.

We are moving very fast. Our plan is to, in the near future, we can already have a product and make the technology so perfect, good enough to have a little product out to the market very soon. We're working with a lot of partners, several major partners such as car manufacturers, logistic companies trying to commercialize our product as soon as possible.

[0:38:40.2] SC: It doesn't sound like the model is one like I think what Kama is trying to do to have like an aftermarket kit that you can just deploy on your vehicle?

[0:38:52.4] JX: Yeah, we are not interested in aftermarket at all, because aftermarket in some sense is almost not doable. There are two aspect that make it not doable. The first is because we need to setup the camera, we need to setup the computer, the modification of the car is a lot, quite a lot.

Then most people simply do not have the skill set, or they just don't want the car look really ugly with a lot of things dangling around. Yeah, it's just not very safe as well. What happen if the camera fall to the ground and then while you're autonomous driving? It's just simply not very easy to do that.

That's also a even more fundamental problem that nobody point is that most vehicle available today on the market already get acquired – already get bought by the customers. They do not support drive by while. That means the steering wheel, the brake, the throttle, the acceleration, they all have to control manually, mechanically.

There's no way, no elegant way, no easy way that you can have a computer to control that for you. If that's the case, most people and they have these aftermarket kit, they'd still cannot control their car. For example, if you look at Commodore AI's modification of the car, if you look at details, you actually cannot control the car under very low speed.

You can only take over the control of the car. For example, about 20 or 25 miles per hour, if you are driving at 10 miles per hour, the computer cannot control it. There's always a certain limitation for those car, for aftermarket will have to fit in.

That's why we are focused on – in our company, we put safety as the primary goal for employment of any autonomous driving technology. So we not only the software have to be smart and not to be safe, but the hardware it has to be good enough to provide redundancy, as well as provide very solid hard work that can run at least a few years, I would say for autonomous driving. I would say don't drive for three days and the camera fall on the bottom, then we'll have enough.

[0:41:03.6] JM: Right. Okay, great. Well, I really enjoyed this discussion and learned a ton about the autonomous vehicle space in general, not to mention what AutoX is doing. Is there anything else that you'd like to share with us?

[0:41:17.5] JX: Yeah, sure. I mentioned, we're a very young company and we're still quickly growing. Right now we have a member of 30 and we're trying to grow to at least a 100 in a year. So we're actively recruiting. If the audience are interested in experience and also working together the cutting-edge of research to determine a difference to the world, please come to find us.

You can visit our website at www.autox.ai to learn more about our opening. Thank you very much.

[0:41:49.1] JM: Awesome. Awesome. Thanks, Jianxiong. I really appreciate it and it was great chatting with you.

[0:41:53.2] JX: Okay. Great. Likewise. Thank you very much.

[END OF INTERVIEW]

[0:42:00.2] JM: 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 Jianxiong or any of the topics covered in this episode, head on over to twimlai.com/talk/58. To follow along with the autonomous vehicle series, visit twimlai.com/ab2017.

Of course, please, please, please send us any questions or comments you may have for us for our guests via Twitter @TWiMLAI, or directly to me @SamCharrington, or leave a comment on the show notes page.

Also, be sure to check out our last show TWiML Talk #57 with Mighty AI co-founder and CEO Daryn Nakhuda at twimlai.com/talk/57 and check out some of the interesting things they're working on at www.mty.ai.

Thanks again for listening and catch you next time.

[END]

