EPISODE 65

[INTRODUCTION]

[0:00:10.7] 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.

I like to I like the start of the show by sending out a huge thank you to everyone listening. We've dropped a ton of great interviews over the past few weeks, and through your dedication we continue to see a growing outpouring of feedback, comments and shares with each release.

If you're a regular listener but don't normally send in feedback, we'd really love to hear from you. Please head on over to Apple podcasts or wherever you listen and leave us a review. A five-star review is of course appreciated, but what's most important is that your voice is heard. It lets us know what you like, or what you feel we can improve on and it also lets those looking for a new machine learning and AI podcast know that they should join the TWiML community.

Speaking of community the details of our next TWiML online meet-up have been posted on Tuesday, November 14th at 3 PM Pacific time, we'll be joined by Kevin T, who will be presenting his paper, Active Preference Learning for Personalized Portfolio Construction.

If you've already registered for the meet-up, you should have received an invitation with all the details. If you still need to register, head on over to twimlai.com/meetup to do so. We hope to see you there.

Now as some of you may know, we spent a few days last week in New York City hosted by our great friends at NYU Future Labs. About six months ago, we covered their Inaugural AI Summit, an event they hosted to showcase the startups in the first batch of their AI NexusLab program, as well as the impressive AI talent in the New York City ecosystem.

Well, we were more than excited when we found out they would be having a second summit so soon. This time, we had the pleasure of interviewing the four startups of the second AI NexusLab batch; Mt. Cleverest, Bite.ai, SecondMind and Bowtie Labs.

We also interviewed a bunch of the speakers from the event and we'll be sharing those discussions over the upcoming weeks. In this episode, you'll hear from Bite.ai, a startup founded by Vinay Anantharaman and Michal Wolski; founders who met working at Clarifai, another NYU Future Labs alumni company, whose CEO, Matt Zeiler I interviewed on TWiML Talk number 22.

Bite is using conversational neural networks and other machine learning to help computers understand and reason about food. Their product is an app called Bitesnap, which provides users with detailed nutritional information about the foods they're about to eat using just a photo and a serving size.

We dive into the details of their app and service the machine learning models and pipeline that enable it and how they plan to compete with other apps targeting dieters and more.

[INTERVIEW]

[0:03:16.8] SC: All right, everyone. I am here at NYU Future Labs in New York City. I am with the co-founders of Bite.ai. In particular I'm with Vinay Anantharaman and Michal Wolski. Guys, welcome to This Week in Machine Learning and AI.

[0:03:32.5] MW: Yeah, thanks for having us.

[0:03:33.6] VA: Yeah, thank you for having us.

[0:03:35.1] SC: Awesome. Awesome. Why don't we get started by having you tell us a little bit about your backgrounds and then we'll get to what the company is up to.

[0:03:42.6] VA: Cool. As I was introduced, my name is Vinay. I work on data infrastructure and for the last 5, 10 years I've been crawling the web trying to turn that into structured data. Currently, what we're working on is building a food intelligence platform that can understand the world's food.

[0:04:00.2] MW: I'm Michal. The two of us met here in this building at Clarifai. We're the first 10 at the least. I got to work with the other 30 years before starting this company. Before joining

Clarifai, I was at Com US studying applied math and computer science, focused on computer vision.

[0:04:15.5] SC: Okay. Awesome. Awesome. You said understanding food. Tell me a little bit more about what that means. I'm assuming since you met at Clarifai, you're doing something visual?

[0:04:23.2] VA: Yes, exactly. To understand the world's food, we need to be able to have examples of what people are eating day-to-day and most ways that we communicate about food these days is with pictures. We started by building an image recognition model that can understand what we're eating, which means that we can take something that we're eating and translate into set of tags and also give nutritional information.

Where we're going with that is actually we'd love to be able to take in other ways we talk about food. It could be the menus or through text and be able to say, "Hey, it's XYZ. It can be made with these ingredients and we want to build a system that can let other people also be able to use the intelligent understanding we have about food to power their food application."

It can be a nutrition tracker in a healthcare setting. It could also be for recipe website. You have a bunch of pictures. They want to know ingredients it is. That's what we're working.

[0:05:15.3] SC: Interesting. Interesting. Now, I don't know if you've come across Hilary Mason. She has a company called Fast Forward Labs here in New York that was actually recently acquired. But they do data science experiments, if you will. That's one of the things they do at least. They experimented with trying to do this, determine nutritional information based on pictures of food.

I think she mentioned this on a podcast I did with her. She said it was an incredibly difficult problem. I think they ended up giving up on it and moving on to something else. What are you doing differently that makes it tractable?

[0:05:50.8] MW: We purchase from a consumer side protocol Bitesnap. It's an app that helps keep track of what they're eating. Kind of like My Fitness Pal alternative using images as input. What we do is every time a user comes in and wants to log in, they'll take a picture of their food, we give them suggestions of what it might be. Users would then correct this, select the correct choice, or pick something else that we didn't perfect. That was the portion size.

Every time someone logs a meal, we'll get training data to improve value. What we do in the app now is if you take a picture of the meal that we haven't seen before, we can't recognize, we allow you to tell us what it is and now then learns to recognize it. Do one shot learning, so next time you come around you can recognize it.

In doing this, we're collecting these data. Users take about three or four pictures a day. We're going to have this human loop system to help us keep improving.

[0:06:35.7] SC: Interesting. How do you measure accuracy?

[0:06:38.4] MW: On the image recognition or on the nutrition?

[0:06:41.1] SC: On the nutrition side, like the full loop.

[0:06:43.5] MW: Portion size is very important. For now we have the users tell us the portion, so we have a prior – There is this big study called NHANES, where the government collected all these data on people how much they eat. We have a distribution of for example fries, how many fries someone might eat for dinner.

We're using that as a prior – as an example of what the portion size might be. We ask the users to correct that. It's still in them to tell us how much, because it's so crucial these days. I mean, just in general to be able to measure food and get them the numbers, you have to be able to tell us how much there is. We're hoping that by getting this data, later on we could actually start using computer vision to predict the portion size.

[0:07:23.4] SC: Okay. I'll definitely be downloading this and playing around with this. I've been experimenting with ketogenic diet, which my fitness pal, like if you're really into it, you're tracking your micronutrients with every meal. It's a pain in the butt to do.

I've experimented with just taking a picture of the things that I eat, and then going back afterwards and then trying to look at it and figure out, and it's hard for me to figure out, okay what the portion size was. But then also what the fat content was. If you go to a restaurant, the fat content in a given meal can vary pretty dramatically. I mean, those are things that you can address just by training data, because in most cases the users don't even know themselves.

[0:08:06.9] MW: We're hoping to use the data we collect from users and starting to getting location into the system. If you know that you go to Shake Shack and you have a burger at Shake Shack, I would see an example of that burger at Shake Shack and so I'll print the information. We'll able to say it's not just a general burger, it's the Shake Shack burger. These are data we get from their menu.

[0:08:25.0] SC: I was going to ask if you were targeting chain types of meals as – it seems like that's easier than –

[0:08:32.4] MW: We haven't gotten to it, but it's on our roadmap in the next few months. Still kind of pulling that information, studies and location. They come from the other signals to improve the accuracy.

[0:08:43.4] SC: Tell me a little about what some of the big challenges have been for you.

[0:08:47.3] MW: I think for us, biggest one is just like you said it's the hard problem that covers initially. People eat all kinds of foods despite millions of combinations. First experience in that matters. Someone comes in, they try it and doesn't work the first time. They might not come back.

For us, making sure we have good coverage of the foods and we have high accuracy and stuff that you – at home, that's not a restaurant and the meal is important. How about for you Vinay?

[0:09:11.4] VA: I would say that for us, because we do have a consumer app, I think marketing is pretty difficult in the consumer space. Our background is a little more technical. It's new for us to be marketing a consumer app. For us, if the more consumers we get, the more data we can get. At least, we've been lucky with the Future Labs. We've been getting advisors and help, but it's still an ongoing process that we're learning how to actually market the app and get it out there.

[0:09:39.4] SC: Interesting. Can you talk a little bit about the pipelines that you've created to process all the data?

[0:09:47.2] VA: Yeah, sure. As we mentioned, we script a bunch of images from that. We basically built our own tools to annotate and clean data. I guess Michal can fill in there.

[0:09:57.7] MW: We built this tool to pooling images, use active learning to help us quickly annotate millions of images. The two of us manage to annotate 3 million labeled images of the different foods.

[0:10:06.7] SC: The two of you met, labeled 3 million? Over how long?

[0:10:10.5] MW: It was about a week. We used clustering and some classifiers to rank and be smart about how we're doing it; done it before at Clarifai. I have an idea of how to handle this scale of a data set. We went from a model of 16 classes to now a bit over 1,500 different foods. Now that that is out, we're getting all these data from users well, and we're starting to chain – I own that data as well.

[0:10:33.9] SC: You said 1,500 different foods. When I think about how many different foods Fitness Pal has in it, it's multiple of that isn't it?

[0:10:43.6] MW: There is this whole foods that you might go get in a grocery or these homemade meals. That's where we focus on right now. Mostly these are basic ingredients. If you take a plate of pasta, of tomatoes, of states, we might say it's pasta, but we'll also give you options for the noodles, the sauce, the cheese on top and break that down.

Just doing and also handle products, just kind of doing the other side of it, it's where you get the millions of different products. For now, we have a barcode feature and we scan the barcode and we'll tell you nutrient information, you can take a picture of it. Now we got an example of the image, we know what it is, so in the future we can start recognizing the products as well.

[0:11:18.4] SC: A combination between the food detection, as well as like a Google look on those. Like I got to find the products, or –

[0:11:24.3] MW: Google goggles, or –

[0:11:25.8] SC: Yeah. Interesting. Are you also allowing them to track the nutritional information over time? Is it take playing that world as well, or just do they take the data and plug into something else?

[0:11:39.3] VA: It's a full tracker logger will give you the breakdowns for the day, for the week and you can see your history over time. I need this.

[0:11:47.9] SC: Yeah. You should go after this keto market, like there is a – I don't know if that's something that's on your radar.

[0:11:53.9] MW: We're actually getting feedback from our users. A lot of them are doing ketogenic diet. As far as decision for that in the future is build it up to start putting communities together. We have users that have all these content and the pictures and we know there are persons doing ketogenic diet. We might be able to connect into other people who are doing keto, help them discover foods, share what they're eating, start recommending other stuff too.

[0:12:13.2] SC: Okay. What are some of the techniques that you're using on the ML and AI side here?

[0:12:18.6] MW: For image recognition, we're just using standard and neural networks, confusion in neural nets. Still use [inaudible 0:12:22.3] might have the solution to [inaudible 0:12:24.5]. For active learning, just simple like logistic regression.

[0:12:29.7] SC: Simple what?

[0:12:30.3] MW: Just using logistic regression to rank the results based on their beddings. Yeah, it's pretty much a lot of neural networks.

[0:12:37.6] SC: Okay. Cool. Where are you in the life cycle, the product is out and on the various app stores, both – you're not going to tell me that you don't support Android support do you?

[0:12:45.5] VA: No. We actually do support iOS and Android. We realized it's incredibly important. We were on both things, because we're using react native, which for a small team it really helps us, because we can have an app that's available for both platforms. In terms of the product, we actually watched an MVP earlier this year and we've been developing that towards product market fit with other people, for with other apps, like My Fitness Pal, people have some expectations so we're really, really close on closing that out.

In terms of the product, we're now actually focusing to start launching an API, so that people can use our technology and other applications.

[0:13:21.2] SC: That's interesting. I think I look for an API for My Fitness Pal. If there was one, it was like you had to submit a form and get approved and talk to their BDP goal and seem like pretty painful.

[0:13:35.0] VA: Yeah. I think that with – because we're taking a little bit of different approach in terms of our app strategy, because we actually are fine with giving people macros and micronutrients. We find that the information is incredibly valuable for people to make decisions about what they're eating. We're offering that up, and we do have plans and thoughts about opening up the API to users themselves so that they can be creative, so that they can view and visualize their data in their own ways.

We've experiment with that, because we can let you export to CSV or JSON. You can pull that in and build your own visualizations. A few people from the self-quantified community are really about it, because they can actually then make their own collages and their own charts and integrate that into their own self-quantified solutions. That's not our focus, but we would love to enable creativity for them around what they're eating, and that's a big part of their life.

[0:14:25.5] SC: Thinking about again where you're coming from with Clarifai, do you see the Bitesnap product as just bootstrapping an API business, or is that the product?

[0:14:40.7] MW: It's the core product for us. But we see that it's a great way to get data to have this human look system that keeps improving. We've got a lot of in my request for an API from customer research firms, from healthcare side, from hospitals, people on diabetic patients. We're looking to open it up to other people. There's a lot of core technology building up to other people using other ways.

[0:15:01.4] SC: Okay. Interesting. Maybe we can spend some time and you can walk me through the annotation framework that you built out for getting through those images? Still that's an impressive feat.

[0:15:12.1] MW: Yeah. The two pretty much we get content from the web, or you have a class policy of the neural network vision before. What you do is you predict the early images. You got some kind of ranking for each class. Let's say for strawberries, we'll take a classifier that's a weak classifier for strawberries, run all of our images through that, get a signal of who might be a strawberry and use that to rank.

We also do visual clustering. We can say for all these images of food, these are the similar looking cluster for images of strawberries that are in the middle of strawberries. Look at that cluster, and for the whole cluster and make a response of yes, this is a strawberry.

Then we have a classifier and use that input to train and improve, so now we can re-rank the results again. Once you get to the high enough accuracy you can say for the rest of them is, because the classifier is 99% accurate, or it says 95% just say whatever the prediction, something that's a called enable.

[0:16:04.9] SC: Got it. When you annotated three million images, like that initial – you look at that cluster of strawberries and say, "Yes, these are strawberries," that might have knocked out a 100,000 images for you or some large number.

[0:16:19.4] MW: Right. If you take the ranking for all the images for let's say the strawberries, you see stuff on the bottom that's definitely not strawberry so you can say the bottom 20% is negative data now. You take the top 10%. If you look at it and see that it looks like positives. You assume that's positive, you'd up to your classifier based on that data and you retrain and re-rank again moving that data away and keep diving into it.

Clustering from is working a little bit also, because now you can still looking at single image of time and you can say for this cluster, all images look similar. They all look like strawberries. Yeah, I'm sure you've seen clustering on the bedding space, so now you can – so they're spying to an image over time for a whole cluster and say, "Yes, this is a strawberry, or not a strawberry." Then have this active learning system, which will pick up the next stuff to annotate.

[0:17:05.8] SC: You annotate your strawberries and then you have all that information for the next thing you're looking at. To what extent does the annotation for strawberries impact bananas if that's what you're doing next?

[0:17:17.9] MW: It doesn't at all. In fresh restart again would be nice. It turns out the neural networks are very good at handling noise. If you have a classified training on the dirty data, there's – 60% probably that the thing that mentioned strawberry has strawberries and then it's – if you train, you're already getting a pretty good signal to recognize those items. You bootstrap with that and keep improving it with the actual annotations.

The other process we'd like to be able to do is segmentation of rebounding boxes. But it's really expensive to get that data on scale. For us when we recognize like thousands of different foods and take us forever and probably tens of millions of dollars to actually annotate at that level.

[0:17:54.9] SC: That's what I envision you were doing. Right now you're identifying images that have a single thing in them and using that to train?

[0:18:01.6] VA: If you have multiple items in an image, or do one class at a time. Do multiple passes over if it has multiple items. Most of the time with stuff, it's like four or five items in play. It would be like soda, burger and fries. In an example like that, we'd do all the fries first. We might be able to pick up a correlation if there's a burger as well. But we do one class at a time and we just like –

[0:18:24.5] SC: When you say you look at the plate and you do all the fries first, so are you talking about when you're training or when you're doing inference?

[0:18:31.8] VA: When we're doing the cleaning annotation.

[0:18:34.5] SC: The cleaning annotation phase.

[0:18:36.7] VA: We're hoping that if the classifier is good, the next time we go around, the burger classifier picks it up, so we see that image again. We will say if the burger prediction is expecting that.

[0:18:47.1] SC: Okay. This iterative process, to what degree is it – have you automated all that into some kind of pipeline, or is still – are you manually kicking of runs to do annotate for object X, or –

[0:19:00.9] MW: We have a batch job to do the initial annotations. We have a system that's called a percent Django of a react frontend that will take those annotations and then have a simple across recognition VM or a bunch of regression to the ranking after. Once you have the initial suggestions for how it goes, it doesn't have to be a neural network, it could be just what data – if you pick up on keywords with strawberries and images on the website and also much strawberry, you can assume it's a weak label as a chance of having a strawberry. It's the index that and in this database you do scans and can go cross by cross. That pretty much is all I need.

[0:19:35.2] SC: Okay. Interesting. How about the crawling part? Were there any challenges involved in that, or is that pretty straightforward?

[0:19:43.0] VA: For the most part, it's just – it's not a challenge to download lots of stuff. What the real challenge is actually getting decent labels. We figured out some techniques to basically, the right sites and the right places that actually decent labels. We actually even took some unlabeled data so that we could actually supplement thee weak labels we have. We can get more examples and that's how we started. It takes a little bit of time to get it right.

[0:20:09.9] SC: Okay. Interesting. What else are you doing that's cool and interesting and that we should know about?

[0:20:14.5] MW: Right now we have an issue with packaged products, you know a model that works well on whole foods, people seem pretty happy with it. But one we got the huge database of barcode products of scans and system data. Now we're also working in being able to use OCR and computer vision to scan our product, be able to index it, be able to brand it as what the health cleans are, what the ingredients are just using computer vision.

Another thing we're playing with is using AR kit and use AR technologies to do the portion size estimation to give an intern working it out. Then an initial idea for this. But the idea this would be take a picture, we got a point cloud, we have these – pick a portion size for an app and now we have a point cloud, we have the image, we can do some segmentation and we have the portion size. Over time, we'll get more data to actually nail that part of the problem.

[0:21:01.5] SC: How does AR kit work? I haven't had a chance to look under the covers.

[0:21:04.9] MW: We have an intern now. For now you can get horizontal planes and if you like go to point cloud. For us, we have the camera open. Now we'll take a picture of it, so start to record to point cloud, so we can send that information and figure out the size of the items.

[0:21:20.8] SC: Interesting. Are you thinking about taking up the next step, or are you project on top of the – use AR kit the way it's designed and project something on top of the plate that says like, don't eat this or put a X over the fries?

[0:21:37.5] VA: Yeah, that would be great. I mean, in terms of AR kit, we're playing with the different user interfaces. One is like what Michal is mentioning is that we could project a cup to help you understand what a 12 ounce cup versus a 16 ounce cup is and put that next to each other, so you can see like, "Oh, yeah. This is pint glass versus this."

There's another mode where maybe when you tell us to do measuring, because it's not perfectly a cup or perfectly a plate, we can allow people to do measurement. I don't know if you've ever played with these measurement apps. They're pretty easy to use. That's how we envision using AR now. I mean, in the future if they were AR glasses, we'd actually love to be able to project the information that we know about food onto the real world, and that's like a friction-free environment.

A, we're recording what you're reading so you don't have to log. Because if you have another passive device absorbing that information it can understand, "Hey, you're eating pizza with burgers today. Maybe tomorrow you should eat something else to feel better maybe."

[0:22:34.4] MW: Also in general, the goal is to get this very passive mode. Right now we still give you suggestions, but for some classes and pretty much like human accuracy, it's common things like burgers and fries. When we predict in for the most part, it is accurate. We want to get to a point where there's pretty much no user interaction, you take a picture or you have some glasses and go by your daily, you're able to measure all the stuff about your diet; tell if there's weaknesses, tell if you have improved. We kind of have this [inaudible 0:22:55.7] where we practically don't have to do any work.

I think what we notice is people tend to eat similar stuff every day. Soon we'd be able to predict what you'll eat before eating. If you're at home, if you go to a coffee shop and you always have a latte at this coffee shop, we should be able to predict that you had it and log it for you without doing anything.

[0:23:13.7] SC: Like using location services, you just walk in a Starbucks, "Should I add pumpkin spice latte to your daily?"

[0:23:19.2] MW: Yeah. Do it before, it's logging to the meal before you.

[0:23:22.3] VA: Yeah. I mean, even actually for other meals like during the week you usually eat XYZ, so why should you have to go and tap stuff? We can just fill that in, so that the meals that really are different are the ones you have to actually log. We can actually build part of that experience already now. We don't need this AR glasses. We can just do that based on location, plus time, day, what meal it is.

[0:23:47.9] SC: You mentioned some of the challenges you're seeing in terms of getting this out to market. What are the top end things, top three things that you feel like you need to be successful based on where you are now?

[0:24:00.2] VA: I would say that we need to have a product market fit, like relative to the other food logging apps. We just can't be having features missing, because people say, "Oh, I'm so used to this. I need it." That's a known known, we know how to do it. It just takes us a little bit of time to get the features right and make them look beautiful. That's another very important thing with consumer apps is that it has to look really good.

[0:24:20.1] MW: So far, it's really high now.

[0:24:22.4] VA: Secondly, the thing is there is a lot of other apps in the space. For us, their incumbents people know them by name brands like My Fitness Pal. For us, it's we have to become feasible and we probably have to come pick on visible through some other – through other channels than just pure search, because most people are searching for those name brands. I think those are our top two challenges in terms of marketing any of the app out there.

[0:24:46.7] MW: On the tech side, data is always – For us, it's always about having data. Right now, for some of the classes we predict we don't have the nutrition data. We leave each other suggestions, so we might be able to predict the item. But because we don't have interesting data tied to it, yeah – like we don't even show that prediction to users.

[0:25:04.3] SC: Did you gather a database of nutrition data by searching and that kind of thing, or by just finding a database?

[0:25:12.0] MW: USD has a very big and detailed data set. It's about a 100,000 different common things people use in America, and it's broken down by ingredients, the common portion sizes, it has the full institutional breakdown. We're using that for now. Now we're working with pulling restaurant data to pulling product data and help us increase that database.

[0:25:32.8] VA: A shout out to the USDA, they do really good work. Their nutrition database is quite complete. What's interesting is actually we learned that other nutrition database is for other countries, actually depend on the USDA database.

They actually have references to the English database and that's long-term and international expansion is a little bit more tougher problem for us than most people. A, we have to convert our labels, so that things that we call – pizza is probably universal, but in the morning pastries people call those things different things and different languages. We have to do all that translation, not only that within nutritional facts. There is regional variations with how things are prepared. Then of course regional dishes, which actually vary by look depending on where you are.

We've been getting a lot of requests from international users, "Hey, can you recognize XYZ for me?" We're like, "We don't even know what that is."

[0:26:22.9] SC: I mean, there are so many challenges there. All kinds of foods that look like one thing on the outside, but they have stuff in inside.

[0:26:31.2] SC: The nice part about having this app, you get to control yourself, is that we get to consider correlations, we get to see eating patterns and these other signals to improve the predictions based on the time that we can say this is probably a coffee, another would say hot chocolate. Using other signals to improve the predictions.

We're hoping that I won't be able to say given and selecting these four ingredients, we can figure out the portion size based on how we see it going.

[0:26:57.1] MW: Yeah. I mean, just basically taking a recipe you know the ratios. This is actually like – this is actually integrated into some of the USDA data. They have some internal equations. Basically from that we realize, "Hey, you can actually take recipes that understand what the ratios are and we can integrate that to things that – other new things that we learn about from restaurant menus or from web recipes," which makes it even easier to log. You just take a picture of this sandwich and probably these things. For most people, that accuracy is great. Because otherwise they have no nutritional information in front of them.

[0:27:28.3] SC: Okay. Awesome. Well Vinay and Michal, thanks so much for taking the time to chat with us. Really appreciate it and really enjoyed learning about your company.

[0:27:37.8] MW: Thank you for your times.

[0:27:38.6] SC: Thanks.

[END OF INTERVIEW]

[0:27:42.1] SC: All right everyone, that's our show for today. Thanks so much for listening and for your continued feedback and support. For more information on Michal, Vinay, Bite.ai or any of the topics covered in this episode, head on over to twimlai.com/talk/65. To follow along with the NYU Future Labs AI Summit Series which will be piping to your favorite podcatcher all week, visit twimlai.com/ainexuslab2.

Of course, you can send along feedback or questions via Twitter to @twimlai, or @samcharrington, or leave a comment right on the show notes page.

Thanks again to NYU Future Labs for their sponsorship of the show. For more information on the AI NexusLab program, visit futurelabs.nyc.

Of course, thanks again for listening and catch you next time.

[END]