

**EPISODE 68****[INTRODUCTION]**

**[0:00:10.5] 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.

Once again, let's start this show by sending some love out to you, the listeners, for your continued support over the last few weeks and months. This community continues to amaze us, continues to grow and to engage with us, which we love to see. We've said it before, but please don't hesitate to reach out to us with any questions, comments, guest or topic requests, or just a friendly hello via any of our various channels.

You can reach us on our Facebook page or Twitter at [twiml.ai](https://twitter.com/twiml.ai). You can reach me directly @samcharrington on Twitter, or you can email us at [team@twimlai.com](mailto:team@twimlai.com).

Speaking of community, please take note, the next TWiML online meet up is coming up soon. On Tuesday, November 14<sup>th</sup> at 3 p.m. Pacific Time, we'll be joined by Kevin Tee, who'll 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 signup, just head on over to [twimlai.com/meetup](https://twimlai.com/meetup) to do so. We hope to see you there.

Now, as you may now, a few ago we spent some time in New York City hosted by our 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 Accelerator Program, as well as the impressive AI talent in the New York City ecosystem.

This time we had the pleasure of interviewing the four startups from the second AI NexusLab batch, Mt. Cleverest, bite.ai, Second Mind and Bowtie Labs. We also interviewed some of the speakers from the event and we're presenting a couple of those interviews to you this week. If you missed any of the shows in the series, visit [twimlai.com/ainexuslab2](https://twimlai.com/ainexuslab2) to get caught up.

In this episode, I speak with Ross Fadely, a New York based AI lead with Insight Data Science. Insight is a really interesting company offering a free seven-week post-doctoral training fellowship helping individuals to bridge the gap between academia and careers in data science, data engineering and AI.

Ross joined me back stage at the Future Labs Summit after leading a machine learning primer for attendees. Our conversation explored some of the knowledge gaps that Insight has identified in folks coming out of academia and how they structured their program to address them. If you find yourself looking to make this transition, you'll definitely want to check out this episode.

Now, on to the show.

[INTERVIEW]

**[0:03:12.9] SC:** All right everyone, I am backstage at the NYU Skirball Center where the Future Labs group is holding the AI Summit, and I am here with Ross Fadley, who is the AI lead at Insight Data. Ross, welcome to this week in machine learning and AI.

**[0:03:29.3] FR:** Hi. Pleasure to be here.

**[0:03:30.2] SC:** It's great to have you on the show. Why don't we get started by having you tell me a little bit about your background and how you got interested and involved in AI?

**[0:03:38.9] FR:** Sure. I have a background in physics, actually. I was an undergrad, I was really passionate about physics. Ended up going and do a Ph.D. focusing on astrophysics. That time really drove me to understand and appreciate data. During that, I was really focused on doing Bayesian machine learning, because we had relatively small amounts of data, but we needed to quantify the uncertainty over our models.

That I kind of just got forced into because of how I went down my academic path, but it made me really excited by what was happening in machine learning and like some of the new emerging technologies on that front.

**[0:04:17.5] SC:** Awesome. I have talked to a ton of folks who came into machine learning and AI by way of physics.

**[0:04:23.2] FR:** Yeah. One of the interesting things that happened when I was a post-doc most recently at NYU, I was jointly appointed with a computer science group with Rob Fergus who won the image net competition while I was there working with him. It's like a really unique opportunity to apply some of these emerging AI, like deep learning models in the context of astronomy, which was really exciting times.

**[0:04:44.2] SC:** Oh, nice. Did he win that before or after Matt Zeiler?

**[0:04:48.2] FR:** He won it with him.

**[0:04:49.8] SC:** With?

**[0:04:49.6] FR:** Yeah.

**[0:04:49.7] SC:** Okay. This is our second year at the Future Labs Summit, and last year — I keep saying year. Steve keeps saying year on stage. It was actually only six months ago. The second time here, and last time I got to interview Matt and hear a little bit about his story. You're in the same group.

**[0:05:09.7] FR:** Yeah. Matt was Ph.D. student with Rob, while I was Rob's post-doc. It was interesting, because I jointly appointed with the physics group, and so we're doing a little bit of crosstalk in terms of like bringing some computer science applications to astronomy, but also leveraging some astrophysical knowledge when applying these techniques.

**[0:05:28.0] SC:** Okay. Awesome. Insight Data. What is Insight Data do?

**[0:05:32.4] FR:** Insight is a professional education company. We run free seven-week fellowship programs to help people transition into careers and data. Yeah, specifically it's focused around

—

**[0:05:43.2] SC:** Did you just say free?

**[0:05:45.1] FR:** Yeah.

**[0:05:45.3] SC:** That's amazing.

**[0:05:45.6] FR:** That was free to the fellows. Historically, it came out as a fact that like there are really smart people in academia and companies needed like data scientists but didn't know how to find them and didn't know that these smart people were there, and they didn't know how to connect. Our founder, Jake Klamka, realized that there is this mismatch, and formed the company.

Of course, academics aren't able to pay a fee for education, because they're usually pretty strapped for cash. We find it really important to make this free for the fellows and the company's sponsor the program, basically.

**[0:06:20.3] SC:** What is the typical fellow profile?

**[0:06:22.9] FR:** It depends on the program. In data science, it's typically a Ph.D. in a quantitative field. It does expand traditional backgrounds, like physics, math, computer science, but it also expands in the biological sciences, even things like political science and linguistics, kind of a mix. That's our data science program, but we also run a data engineering program, a health data program, and I am in charge of the AI program that we run here in New York.

**[0:06:51.5] SC:** Okay. The AI program and the data science program are distinct?

**[0:06:55.1] FR:** Yeah. That's correct. The AI program is a little bit different in terms of background, because we see this mismatch of what people want in the industry and where people are coming from. Often, people in the industry are looking for two broad buckets. One is like the AI engineer, so someone who has extreme engineering talent, maybe knows the machine learning parts quite well, but maybe not to the same degree that someone who's been doing research for a long time.

**[0:07:19.8] SC:** Right. Yeah. Okay. Is the idea that the AI program accommodates both this machine learning engineer and someone who's more traditionally like a data scientist, but focused on AI, or that the data scientist goes through the data science program and the AI program kind of fills the rest of the need for this around this engineering role?

**[0:07:45.5] FR:** Yeah. No, it's more of the former.

**[0:07:47.5] SC:** I thought for sure I was wrong as I was saying that. Why would I do that?

**[0:07:50.2] FR:** No. You nailed it. You're quite right actually. Because like some of these deep learning models are quite specific, quite new and really cutting-edge, the AI program tries to bring in people from that group of research into the program, which is not necessarily true for data science more generally.

**[0:08:07.3] SC:** Okay, got it. The program is kind of like this farm team, feeder program kind of thing specifically for academics. It's not like your typical coding boot camp or something like that where you're taking people off the street and training them up.

**[0:08:23.8] FR:** We actually do admit people from industry, but only for our — Currently, at least, for our data engineering AI programs. Again, on this engineering front for AI for instance, some of the best engineers are people in the industry. We've had people from teams like Google, LinkedIn, Salesforce, top engineers who want to move more into AI come through our programs.

**[0:08:46.2] SC:** Okay. Oh, wow! You talked a little bit about it some of the missing — The things that the industry is looking for that are missing at like a really high level. Can you go a little more deep on that? What are the things that you're really trying to teach these folks as they come to the program?

**[0:09:08.0] FR:** Often, the people who are coming into our program have like 95% to 99% of the technical skills you need. They often come in with an experience, say, deep in computer vision and are well-suited for those types of roles, but they don't know the tradeoffs and how people are actually using this in practice on real teams.

One of the things we do is to help give them this idea of not only like scoping and appropriately you're like doing a project that speaks to the type of work that they're going to be doing in the industry, but we bring in top mentors from teams who tell them about the work they're doing, help them understand what's the important thing and what are the critical challenges they're currently facing. That sort of information is something that you — It's really hard to get isolated on your own in a black box.

**[0:09:54.8] SC:** Is it like project-oriented or is it more traditional just model by model or field by field?

**[0:10:04.5] FR:** Yeah. We find it's really important to not have like course work or book work. It's all project oriented. The fellows execute like significant data science, data engineering AI projects during their time at Insight and it's done in a collaborative way. It's a very collaborative environment. It's fulltime, so everyone is together in a room. Super smart people, and they leverage off each other's knowledge, because one person might be a deep expert in NLP, the other one may know nothing about it. When they started talking they learn much more rapidly than if they are told to like go off in a corner and read a book.

**[0:10:38.7] SC:** Okay. How many people in a given cohort?

**[0:10:41.8] FR:** There's roughly people in each cohort. We're currently in four cities and we run those three times a year. That gives you a little bit of a scale there.

**[0:10:51.9] SC:** What cities?

**[0:10:52.9] FR:** We started out in Palo Alto in 2012, are in New York, Boston and Seattle, and we also run a remote version.

**[0:11:00.6] SC:** I could have guessed those, right?

**[0:11:01.7] FR:** Yeah.

**[0:11:03.5] SC:** Are you expanding to more cities or are you adding more programs? What's the expansion plan, if any?

**[0:11:11.4] FR:** Yeah. That's a great question. To give you a little understanding of our previous expansion, we kicked off New York in 2014 along with a new program in data engineering. We started as just data science. We then launched Boston, focused on health data. That was also in 2014, I believe. Then most recently, Seattle on data science.

Just last year was our first launch of the AI program, which started out from Palo Alto, ran for the first time in New York over the summer. That's like our historical growth. We're planning to move into new cities in the coming future, possibly in international locations and thinking heavily about new programs as they come up in the industry.

The goal really is like to live on this edge of the mismatch between what companies need and are struggling to find good people in and where there's a gap in terms of talent and how can we bridge that gap.

**[0:12:05.2] SC:** It strikes me that you have to address some of these challenges and this kind of impedance mismatch via projects, but also it's also challenging to do so because it's you don't know what you don't know kind of problem. How do you supplement the project work with the things that they need to know, or is it just go bang your head against the wall and figure it out. When you finally figure out what you don't know, we'll be here to answer the question for you.

**[0:12:34.0] FR:** Yeah. As program directors, as people who run these programs, their job is to ensure that they have the resources to overcome technical hurdles as they approach them. We encourage them to first dive into it and hit the walls themselves, but then just basically give them the right tools and guidance to overcome those technical hurdles.

Some of that might come from us, like as people who run the programs, but we have over 1,200 people who've gone through our programs currently and they come back and regularly mentor the people in the program not only on the technical front, but also like what is like to be a data scientist at X-company and so on.

**[0:13:13.4] SC:** Oh, very cool. Is the program run as a single project, kind of like a capstone thing or is it a series of smaller projects?

**[0:13:23.7] FR:** Yeah. It is just a single project, because we want it to have a significant amount of depth that it's like impressive to see what someone can actually accomplish in just three or four weeks' time. Later on the program, the fellows are going around and presenting this work as a way to warm start the interview process with companies that are interested in meeting them.

**[0:13:43.7] SC:** Sure. Are the projects things that they work on individually or as part of teams? Understanding that there's collaboration and people helping each other. Are folks grouped into smaller teams to work on things?

**[0:13:57.0] FR:** Yeah. It's really individual-based, because ultimately these people are hoping to launch their new career. When they're going to meet a new hiring manager and having some of their first interviews, being able to talk about something they concretely did and not having mixed signals of who did what. It's really important.

**[0:14:12.6] SC:** Okay. I guess I was wondering if you had any startups come out of this. If teams, I guess, or maybe individuals, like post-built something as part of this program and decided, "Well. Forget industry. I'm going to — Forget a job. I'm going to start a company around this thing that I did."

**[0:14:31.2] FR:** Yeah. There've been a few examples of that. We're seeing more of it sort of come about especially with the AI program in particular, because there's a new sort of green space that's being created around AI startups. We're working closely with fellows who are interested in launching these new initiatives as well as ones who want something like a more traditional path.

**[0:14:53.0] SC:** Okay. If you could give a list of folks who — Or if you could give folks who are similarly situated in academia or have strong backgrounds in the theory of machine learning and AI, but not the practical experience. If you can give those folks a list of you tell me how long, 1,



3, 5, 10 things to go figure out or learn about in lieu of an experience like Insight. What's on that list?

**[0:15:26.5] FR:** Often, for people, and especially in pure academics, a little bit is about tooling. Some of them might be using a language like MATLAB or R or other languages that may be aren't sort of industry standard. Making sure that you have the tooling to do that. Secondly is sort of like the techniques. Often, when you're doing a Ph.D., you get a book learning of the broad landscape of machine learning, deep learning, and then you dive deeply on one specific model.

Often, when you're on these teams, you want to be able to speak to more broad topics. Not just doing one type of NLP, but knowing the broad spectrum in a more deep way. The way you might be able to do is to actually take on concrete projects with not just clean datasets that come from like Kaggle or some machine learning depository, but actually real live data that either you scraped or you've gotten from like a messy dump of data and dealing with those problems that come up when you're trying to implement something that's more realistic that companies have to deal with.

**[0:16:30.3] SC:** Okay. Two things. Anything — That's consistent with — Yesterday, I was interviewing the startups that are part of the AI NexusLab, and I asked them all what are the biggest challenges they're experiencing, and your later point, dealing with acquiring and annotating and generally dealing with messy data was universally like the number one challenge that they faced. It sounds like it's not just startups. That's what industry folks are telling you as well. Anything else beyond those two?

**[0:17:02.7] FR:** I think a lot of it too is like knowing how data fits in terms of strategy for an organization. I think a lot of people who are academics think, "Oh! You just go and do your machine learning model and check that box." In fact, being able to evaluate how to core the company's mission where work is and what opportunities you're going to have is a very challenging space, and often it will affect how happy you are, how well you fit into a particular job.

On that front, it's more about getting out, trying to find ways to network and learn more from people in the industry and get a sense of like what does it mean to be a big enterprise company who's focused on NLP versus a startup that's doing computer vision stuff? Those are broadly different, motivations and problems that those two organizations have.

**[0:17:49.7] SC:** Right. Awesome. What's next for Insight? We talked a little bit about expansion, but any other programs or things that you'd want folks to know about?

**[0:17:59.2] FR:** Yeah. We're currently thinking of ways that we might be able to work most closely with some of the teams that we work with specifically around like helping them overcome technical hurdles, because the depth of technical knowledge our teams is quite strong. We do a lot of like in-house research in order to stay on top of the type of work that our fellows are doing.

**[0:18:20.4] SC:** Okay. Interesting. How big is that research team?

**[0:18:24.8] FR:** Everyone's sort of like — All the technical team members at Insight do their own like either projects or keeping up with the latest journals. It's not like a strongly unified team. It's more like everyone knows that they want to — They're excited to do it, because frankly almost all the technical team members were people who went through the program. They themselves are very excited by this type of work and are eager to do it.

**[0:18:49.2] SC:** Okay. Awesome. Where can folks learn more and is there a link to the application or what's the process for folks that are interested?

**[0:18:59.4] FR:** Yeah. You can find all the information on our website, [insightdatascience.com](http://insightdatascience.com). There's links to all of our programs including the whitepapers, which actually do a really great job of laying out what is the different between data science. What is the difference between data engineering? What is the different between AI and all those things? A lot of the people, especially from academia and other places, like that is a hard thing to parse and is still like a kind of a hard thing to parse for a lot of people in general.

Great information there, including our blog post, which give you a sense of like the type of work fellows do. Then the applications, we run these programs roughly three times a year. You can

go to the apply page, simple web form, and then there's next steps if the application seems like it's a good fit.

**[0:19:41.0] SC:** Cool. That sounds like an amazing opportunity for folks.

**[0:19:44.2] FR:** Yeah. We think so.

**[0:19:45.5] SC:** Awesome. Thanks so much, Ross, for taking the time to chat with me. I enjoyed learning about the program.

**[0:19:50.5] FR:** Yeah. Thank you.

[END OF INTERVIEW]

**[0:19:56.4] 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 Ross, Insight Data, or any of the topics covered in this episode, head on over to [twimlai.com/talk/68](http://twimlai.com/talk/68).

We hope you enjoyed our NYU Future Labs AI Summit Series. If you need to catch up on any of the episodes, visit [twimlai.com/ainexuslab2](http://twimlai.com/ainexuslab2). Of course, you can send along your feedback or questions via Twitter to @twimlai or @samcharrington or leave a comment or write on the show notes or series pages.

Thanks again to Future Labs for their sponsorship of this series. For more information on the program, visit [futurelabs.nyc](http://futurelabs.nyc). Of course, thank you once again for listening, and catch you next time.

[END]