

Interview with Katherine Luna, Data Scientist at Apple
By Vael Gates, 01/14/2019



Dr. Katherine Luna works as a data scientist at Apple, and kindly agreed to an interview with [Beyond Academia](#), a student-run group aimed at providing resources for PhDs pursuing careers outside academia. Dr. Luna previously earned a Bachelor's degree in Physics and Mathematics from Stanford University, a Master's degree in Financial Mathematics from Stanford, and a PhD in Physics from Stanford. She went on to work at Guardian Analytics as a data algorithms scientist, before arriving at Apple. She cheerfully described her work and job on an early December morning.

--

Vael: What is it that you do?

Dr. Luna: I work at Apple as a data scientist in Search, and in particular, I work on auto-replacements. For example, when you type in “fac” on your phone, it completes to “Facebook” or “Facebook stock breaking news”. And [I work on] how you do the appropriate ranking of more long-term steady state queries, compared to more recent queries that are more trendy, and things like that.

Vael: It looks like you've had at least two data science jobs, right?

Dr. Luna: Yep. I worked at a company called Guardian Analytics that did fraud detection for online banking. That was behavioral analytics, and it was a smaller company with a little under a hundred people.

Vael: How have you found the difference between the two?

Dr. Luna: Smaller startup companies are fun. We had Nerf gun wars, we played foosball together, we had running clubs; it was a very tight-knit community. That was great in terms of— I was doing a lot more statistics-focused stuff just because they didn't have as much data, so you really needed to squeeze all the information that you could out of it. And a lot of times I was talking to customers and putting on a lot of different hats, because at a smaller startup you can't necessarily just focus on research or machine learning, you have to do a lot more stuff. I also played the role of scrum master for a team for a project. [My role was] to drive the success of it, making sure people were on track, doing they wanted to do, what was going to happen next. I had more of a project manager kind of role for that one. There are various product managers, QA people, engineers, front-end and back-end. There are opps people involved to try to get a new product out. It was getting a little boring just doing the same thing of data science in the company,

because we weren't necessarily growing a lot and there wasn't enough more data or diversity coming in. I needed some extra things to do, and I thought, eventually I want to lead a group of say, data science, so I thought that this would be a good way to get into it. Show a little bit of leadership skills, and say I've done certain things where I've managed a team.

Vael: Did you end up doing that transition [to product manager] when you got to Apple?

Dr. Luna: At Apple, I was now getting into bigger data. Before Apple, I didn't do Hadoop. I didn't do MapReduce jobs, Hive, or Spark. I had to learn all that stuff. I wanted to focus more on machine-learning and how to do that at scale, which I hadn't in my smaller company. At Apple I was more interested in doing that, and getting into deep learning as well. When I finished my PhD in 2013, deep learning still wasn't a thing. I took Andrew Ng's [machine learning] course [while I was at Stanford], and maybe just for one day, as a subtopic, we talked about neural networks. That was it, because it wasn't really used in practice. But now it's starting to be used a lot more, so I want to keep up to date with that. I want to learn that stuff first before I take a leadership role doing anything. I think it's important still to learn how to do things.

Vael: It seems like data science can encompass a lot of things. Do you have a good summary, or an example of the diversity?

Dr. Luna: There is a lot of diversity when people say data science. First, I think: are you like a business analyst? Are you more of a machine learning engineer, or are you more of a statistician? There are different flavors of data scientist. For example in Siri, we have people that are more like data analysts or data scientists that work on business metrics. How are we doing in our product? Where do we need to improve? These are business metrics that you would then communicate higher up. How is our platform performing? These are analyzing the health of the product itself, and where you go in the future.

Then you have more machine learning engineers that do the models. The neural network models, random forests, gradient-boosted decision trees... all of those kind of models where you train it, test it, and implement it in production. There's that sort of stuff.

You could have data scientists of various flavors of: what sort of topic do you know? Are you more of a NLP [Natural Language Processing] person or are you into recommendation engines; are you more into health care or security space? Knowing some of this domain-specific knowledge is also useful, because it gives you some sense of where you need to focus your efforts or what things are important when you're first getting into this.

Vael: For the machine learning stuff: that's not analyzing the data, that's implementing actual products?

Dr. Luna: Well, you still have to analyze the data in some way. For example, for completions I have to say: what sort of prefixes give certain completions, and what areas

are we doing good or bad in? How can I develop models to address where we're not doing as good? That sort of feature. Whereas other data analysts might take a higher level view, asking: what domains are we doing good in? Is it Maps queries? Is it news-related queries, entity queries? We work in combination. Maybe they would say, "oh, these are the focus areas that we need to work in," and so then I build models to try to address them too. I can still do the analysis myself, but it saves time if they also do the analysis. It splits up the work a little bit.

Vael: How did you get into this space? Sometimes I see people do bootcamps [(training programs to transition into data science)], and I don't think you did one of those.

Dr. Luna: No, I didn't do a boot camp. But while I was doing my PhD, I was taking classes on the side in statistics and computer science. I remember I was five classes away from either getting a Master's degree in statistics or Master's degree in finance. They say that the dropout degree for physicists is finance— they call them POWs, for Physicists on Wall Street— so I thought okay, I'll do financial mathematics. But I'm from Los Angeles; I don't like the cold or New York or Chicago, where all this mathematical high frequency trading is going on, and it didn't really seem appealing to me. That got me into data science, because it's still about mathematics, using a lot of the statistics properties that I knew. When I was in Andrew Ng's course on machine learning [at Stanford], I did a project in detecting writing styles, to see if people could mask who they were: in a sense, fraud detection. That got me into my job of doing fraud detection for banking [at Guardian Analytics]. It helped that I actually did a project. Insight Data Science Fellowships [, a bootcamp for PhDs transitioning to data science,] want you to do a project because then you can talk about it in the interview and learn all these techniques. I did the project while I was taking the classes at Stanford. A lot of people focus on their PhDs at first and don't do a lot of side classes, so then they'd have to do Insight Data Science Fellowship or something else [to transition to data science].

Vael: So you got hired straight out [of the PhD]?

Dr. Luna: Yes. I did interview at some of these big companies, and I got rejected. I didn't have a Master's degree in certain things, and I was still learning what it was about because I was in a completely different field. People still like to know that you know what you're doing, so getting that initial boost in a smaller company is good, and then going to a place like Apple. Once you've proved yourself a little bit, you have more information under your belt and you can interview better at bigger companies.

Vael: Do you intend to stay at Apple? Where are you hoping to go?

Dr. Luna: I don't know! That's to be determined. I think especially in computer science and data science, especially in the Bay Area, people tend to move around a bit, whether that's within the same company or outside for different companies. The average lifetime is four years or something like that. At bigger companies it's nice that you can move around within the company, moving to a different group. Then you don't have to interview as much outside, which is a little bit stressful. You can form relationships—

while you're at Apple, they get to know you more, so it's an easier process than if you interview for bigger companies outside of Apple, like at Google, wherever. Bigger companies give you a little bit more flexibility in doing different things, whereas at a startup, you can only work on the small amount of data that they give you. There's only this particular project. So then you definitely have to move a bit more to learn more information if you're at a startup.

Vael: When did you start thinking about data science?

Dr. Luna: ...I feel like I had a rough PhD in some sense. I wanted to work with a certain professor; that professor's wife got breast cancer, so then he didn't want to take on any students. I had to find a new group, so I went to experimental physics. I'm still a bit more of a mathematician, so I was taking classes on the side in mathematics and statistics. And I realized, well, I don't want to be in a lab all the time, because— at one point, I listened to a hundred hours of audiobooks in a week because while I'm doing experiments, it's just hand stuff, like soldering and stuff, and I could listen to audiobooks. Whereas if I'm coding, I can't really listen to audiobooks. For me, [coding] is little bit more mentally interesting. I like the feedback that goes on— there's a faster feedback loop in data science, especially when you work with a lot of other people. Because then you get the final result *faster*, if you combine efforts. In physics it was just me doing the experiment, and it would take however long it would take. There were less people to work with me on the project.

And in data science, you can automate things, where in experimental physics, you're doing the manual job. Because at first when you're exploring something, it's not really worth it to automate it, because you don't even know if it's going to work or not. After a few times it's like: "oh, this works! Then I'll automate it," but there's still a lot of manual effort involved. At the end of the day, what I like most about physics is, "once I have the data, what does it mean?" That lent itself nicely to data science, where it's all about the data and what does it mean, and how can I improve things with the information that I'm given.

Vael: How transferable are the skills from the PhD to your current job?

Dr. Luna: I guess it depends on what area you are. I know some PhDs nowadays actually apply machine to what they're doing. I remember going to a talk recently at [the] NeurIPS [Conference] from astrophysics where they were trying to estimate the density of stars from galaxies with gravitational lensing, and they were using neural networks to do that. If you're showing you know what deep learning is, a lot of these skills are transferable. If you're in a lab where you're just doing Matlab and LabVIEW, then not as transferable. That's where you might need to take some extra courses on the side.

Vael: And what are those extra courses? Presumably machine learning and maybe deep learning?

Dr. Luna: Machine learning, yeah. I feel like you probably want some probability course or some statistics course. Whenever I interview at some places, they want to know: do you know what an expectation value is? If you're a beginning data scientist, they tend to ask you some of those textbook questions from intro-level statistics or probability. What is a maximum likelihood estimator, things like that.

Vael: What's your day-to-day like?

Dr. Luna: I wake up whenever I wake up, and I go to work. Sometimes it's at 9am, and sometimes it's at 10am, and then I'll leave at 6, 7pm. It's a nicer work-life balance than academia. I feel less stressed. I have some projects that I work on that I plan out with my manager: this is what I plan to accomplish in this quarter, or this is a milestone that we have. I just work towards those goals. Every week or two I estimate the things I intend to accomplish, and then I update that information. Right now, I'm collecting data, seeing how we're doing from previous runs, seeing how we can improve, adjusting my training data or my testing data and running new models, doing various combinations of those. It's working towards a goal that we have already planned out. It's up to you how you want to manage your day or time—once a week I send my manager an email saying, “this is what I did this week”.

Vael: So you're mostly on your computer. Do you work in teams at all?

Dr. Luna: We do. We have a small team and sometimes we share efforts. For example, one person might do the data collection and the other person might do the modeling part of it. Then some other person might do some model and ask you, “how did you do yours, can I copy some of the data, where did you do it so that I can extract it?” And get to their solution faster. So there's some collaboration there, and also collaboration across groups. Like someone might be doing some trend-detection models and notice that you did some trend-detection models. So they look at your code, and you tell them how you did it and walk them through your code so they can replicate it.

Vael: I'm visualizing you on your computer most of the day, and you talking to people, then going back to your computer?

Dr. Luna: And meetings, too, sometimes. Status update meetings, and planning meetings as well. “What're you going to accomplish in the quarter”, because a lot of things depend on other people, so “who are the dependencies”, and making sure those are good goals.

Vael: Best and worst parts of your job?

Dr. Luna: I like the work-life balance. I like the fact that there is always something to do. My advisor in physics used to say, let the data lead you, and here there's always data. Some of the times when you're in theoretical physics, or you're doing theory of machine learning, my concern was like: what happens if I have no more good ideas? There's not necessarily data there to drive what I'm doing. It's just my own ideas of what I think is interesting. Working close to the product, you actually get to see the impact that you're

making, what's good, what's bad, and you get to keep on iterating and improving on it. I feel like there's always something to improve, and it's just prioritizing what needs to be done. And I like the fact that I get to work with people more. I feel like I'm more of a social person, whereas in physics there wasn't as much collaboration. Here seems more fun, I would say.

The worst parts of my job... well, that's kind of the opposite. It's not as theoretical as what I was used to in physics, right? What is the deeper meaning behind it, how can I prove it? I do know that there's now starting to be some movement of physics in machine learning: using concepts of physics and theory to improve how machine learning is done. There's some people that I know at Facebook FAIR or Google Brain that are working on such things. For example, symmetry arguments towards improving how graphical models are done, or visual ones, or also how to improve how training is done in terms of neural network models. Like, there are fluctuation dissipation theorems in physics that you could be using that happen in steady state solutions. Neural networks eventually go to steady state, so you can apply certain physics concepts to them to help improve how training is done. That was interesting to see again. It sparks the physicist in me and seems interesting, but at the same time, after a while I'm like, "well, if I keep on doing that..." In the short term that seems interesting, but in the long-term I realized: I kind of want to be somewhat more useful. I feel like, what's the point? At least when I'm doing applied sciences, I get to see the impact, I get to see people enjoying the product more. While one can still do certain things at the end of day, what would make you happy in the end? Everyone's capable of doing various aspects, but at the end it's what makes you happy.

Vael: My next question was going to be what motivates you to do the job you do. I'm hearing something like, it's being able to see the impact of your work?

Dr. Luna: Yeah, be able to see the impact of your work. Also, if you're in applied research science, you get to read other papers and implement them. Maybe you might modify them in certain ways to apply to your particular test, and at the same time maybe you might come up with some new model for which you can write a paper. It's still being at the cutting edge, but also being useful in improving things.

Vael: So you're publishing papers! That's cool.

Dr. Luna: Well, not publishing papers yet. I think that Apple is trying to improve that—where people can have more time to publish papers and more time for research. It's something that is definitely a focus Apple is trying to improve. They recently hired John Giannandrea, who led the AI group at Google. So you want to improve the culture of Apple to be a little bit more research-focused too. Right now it's top down: this needs to get done. You know how to improve it that way, but then it'd also be nice to do a little more research. There's a lot of easy things to do to improve something.

Vael: Just to be clear, your job right now is applied research?

Dr. Luna: Yeah, it's applied research, because you see some of the techniques. You read papers, like "how do other people approach this problem", and then you implement it. I'm a hybrid. I'm like: am I a machine learning engineer, am I a data scientist, am I an applied research scientist? I don't know! I just kind of improve the system.

Vael: Are there any other good motivating factors for like why you get up and think: yep, I'm going to work today! Sounds great.

Dr. Luna: Well! Apple Park is a very nice campus. It's very pretty. I have gelato which is on the same floor as me, which is very tempting all the time. You know, there's something interesting to do during the day, nice people. I think it's important to have a good manager and a nice community, because that makes me happy! As long as I'm around people that are happy and friendly, and I'm doing something that's somewhat interesting, then I'm happy. I'm at a good state, at least.

Vael: Do you have any other knowledge on data science careers in general? Recommendations or knowledge that surprised you when you into the field, anything like that?

Dr. Luna: A wave of schools are now giving Masters [degrees] for data science or machine learning. I've seen that increasing. I think M.I.T. has a program now, Northwestern or Northeastern has a program, Berkeley has a program. That's interesting. I wonder, will that make it harder for people with a PhD to say, "I'm going to switch into data science"? Then will they have to do the Insight Data Science Fellowship for example to gain the skills necessary, or will they have to get a Masters eventually like other people? It's interesting.

But other insights? I feel like when people interview, it's good to have a project, whether they do stuff on like Kaggle, or take classes on Coursera. People like to see that you have a project, you can discuss about it in certain detail, you have a little bit of breadth. I think all of those things are important for interviewing. Normally interviews involve a coding section; they want to make sure you know how to code pretty well. Then there's machine learning kinds of questions. Do you know the difference between naïve Bayes or random forests? How would you train a model from scratch, like the zero-shot model? How do you know if your data model's overtrained or undertrained, or overfitted or underfitted? There's a data science component usually for interviews, and then a product sense as well. For example, when you interview at Facebook, they might say: here's Messenger app. We're thinking about including a new feature. How would you analyze the success of the feature? What kind of metrics would you use?