My Story in Computing

featuring Dr. Eliezer Kanal as Interviewed by Suzanne Miller

Welcome to the SEI Podcast Series. A production of the Carnegie Mellon University Software and Engineering Institute. The SEI is a federally funded research and development center sponsored by the US Department of Defense. A transcript of today's podcast is posted on the SEI website at sei.cmu.edu/podcast.

Suzanne Miller: Hello. My name is Suzanne Miller. I am a principal researcher here at the SEI in the Agile in Government team. Today I am here with my colleague and friend Dr. Eliezer Kanal. We know him as "Elli" because otherwise I can't pronounce his name all the time. He is going to talk to us about his path to getting to the SEI. I am very excited because he and I share some background, and we have different paths of getting here. We want to hear about his path of getting from his early work in *not* software engineering to how he ended up here at the SEI and what kind of work he does today.

Welcome, Elli. I am really excited about today's podcast.

Eliezer Kanal: Thank you. Looking forward to it.

Suzanne: Let's just start off with letting people know, what is it you do here? What kind of work do you do here at the SEI?

Eliezer: My main responsibility is I am in charge of the CERT Data Science Group. The way I describe what we are doing, we're doing phone screens. I've been doing a lot of these lately. In the SEI, this FFRDC [federally funded research and development center], we have quite a number of teams who research all things related to software engineering, cybersecurity. Our team acts for the most part as internal consultants. We work in CERT, so most of our work is in CERT. If I have someone working on malware, we may help them build a classifier for malware. If I have someone who is working with network security, we may help them build some sort of a classifier to identify issues of network security. We will work also across directorates. Within the SSD [Software Solutions Division], there are some folks who are working on network architecture.

We are actually building classifiers to help identify patterns and violations for people who are trying to build software, and it is not following the design pattern that was intended.

Suzanne: That's a little bit different. Data science is not just about the classifiers, but it is also about dealing with big data and dealing with all of the elements of data management as well. There is a lot of different aspects to data science that you guys have got to have some knowledge about.

Eliezer: That is a good point. One of the things we consult on frequently, both internally and externally, [is] how do you handle large data? If you have large sets of data, what kind of questions can you ask? What is the process to realize the questions you have to ask? Where can you put it? How can you actually stick it in a space that's analyzable? What formats do you use? Additionally, just training. We do an awful lot of just [asking], how can I take my team and convert them from whatever they are into data scientists? There is a lot of tangential stuff that goes on here, as well as a lot of just brainstorming, and how do you design a good research project?

Suzanne: We all have to be ready for that. Your impact is mostly on the internal projects. Can you tell us a little bit about any projects where you are particularly proud of, *Because we were able to help them they were able to do this*?

Eliezer: There have been a couple. One of the ones I think was most interesting—the work wrapped up right before I got here three years ago—There is a department in, I guess now under DHS, called US-CERT. One of their jobs is to collect and catalogue all these different incident reports across the government and industry and really keep track of, *When software goes bad, what do you have*? They have hundreds of millions of these tickets just describing all sorts of problems, all sorts of risks, little events that occur.

When you have that much stuff it is kind of impossible to tell what is in it. We worked with them, and with some other teams in SEI, to create a tool that will allow them to visualize, choose a signal data point. It spreads out from that data point searching across all these other tickets and will build you this graph of, *Here is what you found. In this one ticket you found*, maybe we have an IP address, maybe we have a hash, maybe we have a date, a file name, a path name, stuff like that. *Here are all the tickets that have that information*. Maybe these two tickets share a hash, but the second ticket actually has a file name that wasn't in the first. *Let's search around to see what other tickets have that*. You build this tree so to speak that starts to build a picture of what a full attack might look like. Let's say some piece of malware comes in somewhere. I only happen to see one random slice of what it did. By putting together all the relevant tickets, you get to see this whole story of what the malware was doing. The reason I find that so neat is because given the number of tickets, the chance that any one analyst will actually find all the relevant stuff is

zero. It really rounds down to zero. There is no way they can find everything. That is a pretty cool story. There is more if you want.

Suzanne: No. That's the kind of thing…Part of what we want is to help our audience understand why what you do is cool.

Eliezer: Yes. Pretty cool stuff at that.

Suzanne: Being able to go from a single data point to a graph of a potential threat map is pretty cool I think.

Let's talk about how did you get here. You have a premed, psychology, so psychology was the actual degree. You eventually earned a doctorate in biomedical engineering from our local university we are very proud of, University of Pittsburgh. What was your original plan? Where did you think you were going, and I know that's not where you ended up, but where did you think you were going at that point?

Eliezer: When I was growing up, like any good Jewish boy, *You are going to be a doctor, lawyer, or Indian chief* as my dad always says. I grew up thinking, *I am going to be a doctor.* My dad's a doctor. Probably about 9 out of 10 for all my friends, their parents were some form of professional whether it be a doctor, a lawyer, but doctor, lawyer really covers all the bases there. I was planning to go to med school. Somewhere along the way, I decided that psychology was an easier degree to get a better GPA in. So, boom, there you go. Solved that.

As I was going through that process and applying to schools, turns out I wasn't really a great student. I could learn all this stuff, but I just didn't test well. I am going through this process. I am like, *This med school thing may have been a great idea, but it's probably not going to work for me*. I started applying elsewhere, at which point I realized the psychology degree might not have been the best choice. At that point, I tried to say, *Well, I have time to pivot*. I took some classes in the [neuroscience] department. I never actually took a course in programming, but some of the math that I had missed, the differential equations…And I started moving more towards an engineering field.

Suzanne: And the biomed sort of connected in?

Eliezer: The biomed was as I was looking around and seeing what was happening, and I wasn't getting into med school, I saw that University of Pittsburgh had a, at that point, fairly new bio-medical engineering program. They have actually since done me the wonderful favor of going on to becoming a very highly ranked department across the country, which does wonders for me. At the time it was fairly new, I think only six years old when I actually joined. I said, *This sounds fascinating*. I didn't really know too much about it, but I looked at the different professors who

were there, and I ended up joining the lab of Dr. Robert Sclabassi. He was a neurophysiologist in the Children's Hospital. What that practically means is he is looking at the EEGs, electroencephalograms of mostly children as they are going through surgery and making sure that they are staying healthy mentally and physiologically during the surgery. That was most of the work I saw him do. I am sure there was more to it. I joined his lab, and I was doing research at the time on prefrontal cortex functioning.

Suzanne: So decision-making and all of those very intense cognitive things, which has nothing at all to do with engineering software.

Eliezer: Well, no. Actually, it's interesting to tie that back, I was doing work on addiction. I was doing work on learning. I was doing an awful lot of work on memory. When I say work, what I mean is I am reading just paper after paper, after paper, and going through taking all these courses in systems neuroscience and cellular and molecular biology and fun stuff like that. As I am walking through all this and learning this, I start to realize the work that I am doing is a lot of recording brain activity. When you are recording brain activity, you end up with all these time series, lots of time series. I was working in this thing called the Magnetoencephalography, MEG. We got from that 302 sensors recorded at a 1,000 Hertz, so it's an enormous dataset. One hour of recording was a couple gigabytes. Then turning that into something useful was a lot of signal processing. I started taking signal processing courses, and along the way I realized that biology is fun and exciting, but it's an awful, at least for me, it was an awful of remembering things, and I have a terrible memory. Trying to force myself to just remember all these different parts of the brain and how all these different chemicals work and all of the different neurotransmitters, all the different regions, interactions; it was a lot. I really enjoyed though the signal processing aspect of it. That was just where I found it fascinating.

Suzanne: How did you translate from that sort of academic background into the work environment and actually sort of moving towards where you ended up now. It makes perfect sense to me now, how you ended up in data science because that is really what signal processing is about is taking all of the data signals and making sense of them.

Eliezer: I actually started getting into it while I was still in grad school to a certain extent. One of the really hot fields when I was in grad school was this thing called neuroprosthesis where you would take a set of electrodes, place it on the cortex—usually they do this with someone while they're undergoing surgery so the brain's already open—but sometimes they would do it just put a cap on someone's head, and then use gel and stuff to get a signal. They use them to control a mouse or to control a robotic arm, control a wheel chair. There was a whole field that was really, really coming into its own in 2007, 2010.

I was watching that, and I was seeing there were so many companies being brought up around this field of neural engineering. A lot of them were not doing so hot because it was really early technology, but the idea of bringing this into industry was right there. When I finished doing all this work, then I was like, *Well, now, I don't want to go into the bio field*. If you are doing biomedical, almost everyone is going to go into device engineering. *I am going to make the next EKG, make the next scanner...*

Suzanne: Pacemaker...

Eliezer: I didn't have any hardware knowledge whatsoever. A lot of the kids in my program came from an electrical engineering background. A lot of kids in my lab were doing that, and they were actually soldering circuits. I didn't. Because of that, I ended up selling myself as a statistician. I actually worked at the bank, at PNC Bank, for a little while building quantitative models of risk. It was interesting. It's basically saying, *Let's take all this statistical knowledge that I gained doing this PhD, and let's see if I can apply it to do something useful*.

Suzanne: Very practical.

Eliezer: Very practical. It was an interesting experience there.

Suzanne: So you went from neuroscience to banking. Where did you go after banking?

Eliezer: I worked in the bank for a little bit. For what it's worth, almost all of my colleagues had done almost the exact same things as me. A couple of them were from CMU Stats department. They were moving around, but that is a pretty common transition. After the bank, I went to Highmark, and I worked as a manager of metrics and analytics. Actually, the title was Manager of Analytics Application Development. It was mostly managing all the metrics for the operations side of Highmark, the insurance arm. That was pretty nifty. That was my first foray into, *I have to now take an enormous amount of data*—I mean Highmark has billions of claim line items every year—*and trying to take this data and condense it down and format that the business can understand what do they have what are they dealing with*. Question involved things like, *How many adjustments do we have? What kind of adjustments are causing problems? Do we need to adjust our claims processing?* Those sorts of things. Trying to convey that into a way that people actually understood it was the challenge. I really enjoyed that, even just building a dashboard. It takes an awful lot of insight to build something that people can look at and actually understand.

Suzanne: Especially if you know how much data's behind it, that's the thing. The dashboard is, I call it apparently simple and inherently complex because you have to make it look simple, but you have got all this complexity behind it that you are managing behind the scenes and trying to make sure it's accurate, and trying to make sure you don't misrepresent. All those things are things that people just don't think about when they see the dashboard when it is completed.

Eliezer: Very much so. Every data scientist eventually comes to learn that the algorithms part of being a data scientist/machine-learning/AI expert is maybe 5 percent of the job, 90 percent of the job is just wrangling data. So especially when you are working with the business people in such an applied industry. Getting through all the nuances, *Well you know you did this [data] pull but you didn't exclude that group and this particular group had a problem on that one day. Did you exclude that because otherwise…* All these little nuances come to bite you. I had this one guy on the team, Adam, who was unbelievable at making this stuff…another guy, Steve, a whole bunch of folks who were pulling all of these things together and did an absolutely fantastic job of really putting this stuff in a way that the business could understand.

Suzanne: This comes back to one of the things that a lot of our work revolves around, metadata and around, *How do we characterize the data itself*? We are characterizing information and synthesizing and abstracting. *How do you do that in a way that makes it useful for the purpose, also, in our case, useful for other kinds of research?* We have a lot of different purposes to the data that we use. You have got the real in depth look at, what are the different ways a business can want data to be managed and manipulated so that they get the answers that they need.

Eliezer: Very much so and even more so, the business was not…I used to say when I was working at the bank, *PNC is actually a data company who runs a banking business*. No one listened to me at the time. At PNC, I was a nobody. I was one of the thousands of analysts they had. At Highmark I had a little bit more of a voice, and I was starting to convince them, *You guys are a data company that happens to sell insurance*. I was able to convey that through piece by piece trying to show, *Look at what you can do with what you have*. I was there for a little bit and I got to know…They had an informatics team whose job…My job was operations. *Are the phones up today? How many ID cards didn't get sent out?* Stuff like that. The informatics team...

Suzanne: The medical informatics...

Eliezer: ...their goal was what information is hidden in this data. We would put together all sorts of fascinating stuff describing populations and trying to find areas at risk, people who they are undeserving. *Maybe we need to optimize our phone tree*? There is all sorts of information that you can pull out and conveying that mindset of, *You have such a rich trove of data. All it takes is the right question and the right approach to just, How do I deal with this. How much value do I have to be extracted*? That was a ton of fun.

Suzanne: *How far does it have to be synthesized before it has meaning*?

Eliezer: Honestly, that was my job. That was my job. I was just trying to convey to the business the concept of, *Ask questions, and know that you have the data*. *Now that you know this data can be kind of questioned, think of the questions you want to ask*. Then come to me, and I'll tell you,

All right. Now we have to do X, Y and Z. Or, *This one's tougher to answer than others*. Just think about the data as something valuable you own.

Suzanne: A valuable asset.

Eliezer: A valuable asset. Rather than saying, *Oh I have membership, and I'll use that to query whether the guy is a member*. *No I can use this membership data to learn about my population, to learn about the customers: insights about where I should be growing or how I can compete better*. That sort of stuff. It was a lot of fun to hand over.

Suzanne: What was it that made you come out of business into the SEI into this research environment that we have?

Eliezer: The SEI has something really unique, which is CMU. The SEI has something else really unique, which is government. As I tell everyone, this place is one of the best places to work because on the one hand, I can look out my window and I'm staring at the CS [School of Computer Science] department. I can take courses in the CS department if I want. I can partner with some of the smartest people that are in the area. On the other hand, I can also go down to work with some of America's finest working in the military. You can go onto the Hill and try to talk to people and convince them that there are certain policy changes should be made. We happen to have even on campus some people who have been formally in politics and therefore, they have connections. If we have thoughts about hey, maybe we have some ideas about how we should be doing cybersecurity policy in the government, we actually have a voice and we have some power behind us because of who we are. The CMU and the SEI name carry some weight. It's a fantastically fun place. When you are looking at Highmark, it's a good company and there's a lot of stuff going on for it. I think the impact here has a lot more pressure.

Suzanne: A lot of us come here for the impact. I agree with you: the uniqueness of having CMU as the academic arm. It's not a lower-tier university. It's a top-tier university. So the people you get to talk to really do know their stuff. Then being able to translate that into something useful for our government customers. I know that's a big reason that I am here.

Eliezer: I'll tell you even more so. For some reason I didn't expect this. The faculty are really fun to work with. You expect them to be some sort of stuck up something. No. these people are great. I have been working with these different professors. They are more than happy to talk to you. They are down to Earth. They are like *Oh, good idea, good discussion*. It's really heartening.

Suzanne: I agree with you. There is a lot of intellectual energy, I guess is the way I'd put it, and everybody is curious. Working with people who are curious just is a joy.

Eliezer: Intellectual energy then translates into impact. CMU has an over-large focus on societal impact for an academic institution, which is wonderful. I find that very nice.

Suzanne: If you are talking to teenagers today who are thinking about their future, what would you tell them about, why is it exciting to think about working in this area. What would you tell them about their education? You certainly did not have a traditional computer science education. What would you tell them?

Eliezer: I never took a single course in programming. What I would tell them is learn to program. I don't care whether you're planning to go into hairstyling or acting or whatever, but learn to program. There will be immense value, no what you do, in just knowing how to tell a computer what to do and then making it do it for you.

Suzanne: That's all a computer can do is what you tell it. That's what I tell kids. If you don't know how to tell the computer what to do, it's not going to know what to do. Whatever it is you are doing, you are probably going to have to deal with a computer.

Eliezer: Exactly. Just learn to program. Even better now there is so many languages out there that are specifically intended for kids. People think, *Oh I need to learn to program, I need to learn to type all this stuff*. No, there's a program out there called Scratch. I am more than happy to sell it. It comes from MIT. It is basically Legos. You just drag and drop, little click, little pieces together, very intuitive. There are wonderful videos out there how to do it. You don't have to type a single letter, and you can make the computer do…My kids have written coloring books. They have made games. There's all sorts of fun stuff you can do with just this simple interface.

Then people say, *Well that's not what programming is like*. Not all, when I was doing my PhD, one of the more powerful pieces of academic software is this thing called LabVIEW. It's not even academic it's used in industry. It's drag and drop just as well. You can build neural implants using LabVIEW, which is just Legos for programmers. I mean learn how to use these things. These tools are out there, and it's really, really worthwhile.

Once you have learned to program, I would definitely say at least take a peek at some of the science and analytics. I mean the people who know statistics know how to interpret data. The benefit they have over everyone else is immeasurable because you understand how to interpret data. Every time you open up the newspaper—if anyone does that anymore—read a blog online, watch television show, someone's trying to convince you of some fact. If you don't have an analytic training, and if you don't know, *Is this data is collected correctly? Was the survey put out correctly? Are they leaving out some enormous biases*? If you haven't been introduced to that kind of thinking, you are basically either forced to say, *I am not going to listen to this*

because I don't know whether it's true or just swallow it whole. But having analytics training gives you the ability not to do that and lets you understand how to work with data.

Suzanne: There was an incident now too long ago when I got one of these random calls for a political survey. There was this election coming up. It took about 20 minutes. It took way too long, but it was interesting to me as an analyst. I am listening, *Oh, that is the third time they have asked that question a different way. Oh look, they are trying to do the reverse*. Some of it I could say, *Oh, this is good. They are trying to get to see if there is bias. They are trying to make sure that I answer the no question no and the yes question yes to make sure that's what I meant and things like that.* There were other things where you've just like, *Oh come on. You're trying to bias the data for the result that you want*. If you don't have a sense of that, you can end up being duped for all kinds of purposes that you really don't intend.

Eliezer: Very much so, especially if you've seen certain organizations like to send out mailings. Choose an issue, choose an aisle and whatever on both side, they'll send out mailings with surveys and you can clearly tell these things are driven one way or the other. That is some of the more obvious stuff. As you said, there are many opportunities for people who don't know basic, I'll say research methods but also it's just understanding how to analyze stuff. You might not notice it, and you really want to be cognizant of this stuff.

Suzanne: Where did you get your career advice? What helped you to make the decisions along your path, because none of us do this without external influence? Where did you get the best advice for helping you move your career forward?

Eliezer: It is interesting. The best career advice I got was unfortunately way too far into my career, but I'll say it anyway. Once I was already at Highmark, and I was a brand new manager, and I was just learning how to work with a team. I was talking to a guy, this is a manager that I'd known from beforehand, a wonderful guy, Scott. He told me after one of my first phone calls with a couple of muck-de-mucks, he pulled me aside and he's like, *Elli, when you are doing anything, your team absorbs all the credit, and you absorb all the credit and you absorb all the blame*. I have tried to extend that out more so like whenever I am doing anything, making sure that you have appropriate interactions with the folks you are working with. Always be humble about what your own contributions were and brag about...If you're the one talking Clearly, you are the one who did it and everyone knows that because if you didn't you wouldn't be the one talking. You don't need to brag about your own part. But be generous with what you give out. That's a bit more specific to team management.

Regarding the career stuff, to what you said, everyone goes throw the stuff that they have picked up. A lot of it just comes from watching the colleagues that I have and sort of picking up from them, these are areas you really want to be active in. As I said before, this sort of knowing how

to use a computer well, knowing how to understand numbers well and these are things that I have seen that are super critical.

Suzanne: As you said, they don't necessarily come from just that one path. Psychology, you get a lot of statistics, Biology, neurology, you've got a huge dataset. There's lots of different paths that people can come to engineering-ish kinds of activities.

Eliezer: I am in a position, a little while now, where I have been interviewing for this job and other positions. I do interviews from people coming to the jobs that I am posting. We get people from any which where. Some of them come with CS degrees, some come with stats degrees. A number of them come with degrees like I have from the neurosciences. We have had at least, actually quite a few people coming with astronomy degrees, because they're...

Suzanne: Wow. That is also big data.

Eliezer: …enormous amounts of data. We had a few folks come with history degrees. I know of at least one person who came with a theatre degree. There is all sorts out there. A lot of people come in with bio. Let's say you are doing wet lab work not computational bio, *What can I do here*? As I said, assuming these guys know how to use a computer because you have to analyze some data, these are people who know how to do research methods. Once you know how to do basic stats, you can work through a lot.

Suzanne: They know how to deal with complex systems, because one of the things you and I said before we got started, biological systems are way more complex than any mechanical system that we can conceive of. I actually like working with bio background people because I know that they have got some things in their head to help them deal with the level of complexity and partitioning and some of the things that we need in computer science. They don't even know they know it but it's what they've had to learn to be able to deal with the complexity of that discipline.

Eliezer: Coming from neuroscience for example, there are people who study literally parts of the cell wall of a neuron, so one or two actual junctions in the neural wall. Then there are people who study the cell. Then there are people who study groups of cells, then tissues, then systems of tissues. Then you have the entire visual cortex. Then you have chains through, *How do I recognize a voice*? It's kind of engrained at least when I was doing this at Pitt, it's engrained in you, You are going to step from these levels. The cut is not always so clear, but you have to appreciate how much complexity you are washing away at each level. It's fascinating, exactly what you're saying. There's so much we don't understand, but even within that, let's try to isolate it down into a reasonable experiment that we can find something useful.

Suzanne: I want to thank you for having this conversation with me today. I am so glad I got to be the one to interview you because talking about neuroscience and biology takes me back to my own past, so this was quite wonderful.

Eliezer: Thank you.

Suzanne: I hoped that it helped some of our audience understand that you can get to pretty much anywhere in the analytics world from almost anywhere that you come from.

Eliezer: Yes, if there is one thing I want to push is really that point. No matter what you're doing, no matter where you are literally in your career. You could be post PhD for English literature, there are people who have made the jump into data. Never think it's impossible. Really, feel free to reach out and talk.

Suzanne: Elli does have a lot of publications in this area. We will provide links to some of those in the transcript so you can get a little bit better feel for exactly he writes about. I do want to thank Elli, and I want to thank our audience today. I hope that you found this useful. We will be having more of these. This is the first in our series on the unique ways that people come to the SEI and to engineering. So, we look forward to us sharing more of those stories with you.

Thanks for joining us. This episode is available where you download podcasts including SoundCloud, Stitcher, TuneIn Radio, Google Podcasts, and Apple Podcasts. It is also available on the SEI website at sei.cmu.edu/podcasts and the SEI's YouTube channel. As always, if you have any questions, please don't hesitate to email us at info@sei.cmu.edu.