

SDS PODCAST

EPISODE 461:

MLOPS FOR

RENEWABLE

ENERGY



- Jon Krohn: This is episode number 461 with Dr. Sam Hinton, astrophysicist and expert in machine learning operations.
- Jon Krohn: Welcome to the SuperDataScience podcast. My name is Jon Krohn, a chief data scientist and best selling author on deep learning. Each week, we bring you inspiring people and ideas to help you build a successful career in data science. Thanks for being here today. And now, let's make the complex simple.
- Jon Krohn: Welcome back to the SuperDataScience Podcast. I'm your host, Jon Krohn. And oh my goodness, are you in for a treat this episode because we're joined by the brilliant polymath Samuel Hinton. Sam is a prominent astrophysicist, a software developer and a former contestant on the television show Survivor, whom I'm sad to say was metaphorically stabbed in the back and voted off the island. But most importantly for this audience, Sam is a data scientist with a particular specialization in machine learning operations. More specifically, he works at a British firm called the Arenko Group, deploying fascinating machine learning models into production systems that enable renewable energy sources such as solar and wind to undercut the pricing of dirtier energy sources like coal and gas.
- Jon Krohn: The first half of today's episode should be of interest to anyone who likes to have their mind blown. In the first half, Sam provides elegant digestible explanations of his astrophysics research, such as why our universe is expanding at an accelerating pace, and how machine learning is used at a terabyte scale to try to understand this phenomenon. The second half of the

episode caters more to hands on data scientists and software engineers who are keen to learn about MLOps, machine learning operations, why we desperately need it if we're deploying machine learning models into production systems, and Sam's top tips for existing tools and techniques in the MLOps space.

Jon Krohn: Before we dig into that, a quick announcement that starting in Episode 465 in two weeks' time, we will begin releasing guest episodes on Tuesday mornings New York time. Historically, we've released Wednesday evenings, but by releasing 36 hours earlier, we'll be giving you two more morning commutes in your week to enjoy the episode. I can't imagine any downsides to this change, but I didn't want to catch you off guard when it happens.

Jon Krohn: Sam, welcome back to the program. It is such a joy to have you here. You were here for Episode 303. You did p-hacking and stats. You were also here for Episode 367 talking about COVID-19 data pipelines. I'm looking forward to some juicy updates on that. But in general, how's life? How have you been since that episode?

Sam Hinton: Oh, I mean, I think I've been as good as everyone can be. We all know how 2020 and 2021's going. So I'm still kicking. And for that, I think that's as good as you should hope for and from that, yeah, I'm great, Jon. How are you?

Jon Krohn: Nice, well, I'm great. Thank you. Thank you for asking. So in that time, you've had some big life changes. You moved from Australia to the UK. Have you ever lived in the UK before? How's it going?

- Sam Hinton: Oh, I have visited the UK a bit. I've worked with people in South Bend and Portsmouth, Cambridge and Oxford. So I've done that too. But I've never actually lived in the UK. So that's definitely been new. I found out a few interesting things. For example, they have Council Tax here that you pay on top of your rent. That was a great surprise. But yeah, since I was last here, uprooted myself from blissful COVID free Australia, decided to travel right into the pit of doom London and see how that goes. And yeah, exciting times.
- Jon Krohn: So you're paying your television tax as well?
- Sam Hinton: What's that television? Is that, oh you mean the Netflix tax?
- Jon Krohn: Yeah, that must be, the internet must be really hemorrhaging Her Majesty's revenue from the television tax. Who doesn't want to live in the UK? We're now making, I'm not making this up. There's a TV tax. It's not a huge amount of money. It's something like 150 or 200 pounds a year or something like that. And you can opt out. But they can come and check and see if you have a TV I guess. I don't know how often that happens. And yeah, in today's world where you don't really need a TV, you can just stream Netflix, you really don't need to pay the television tax.
- Sam Hinton: Is that money that just goes to the BBC for the public funded broadcasters? Because I can sort of get behind that. They do a lot of niche work that people in Hollywood wouldn't do, because there's no money in it. But it seems a bit weird. It's separate from every other



source of tax and how the government works. But the UK is a very weird place, as I've found out.

- Jon Krohn: I think a good chunk of it does, in fact, go to the British Broadcasting Corporation, the BBC, and I think that they do do great work. They do a lot of investigative work. I think they can be trusted. And as someone who grew up in Canada with the Canadian Broadcasting Corporation, I also really appreciate their reporting. And I think a trustworthy public broadcaster could be hugely valuable in the United States where I live now.
- Sam Hinton: Oh, yes. Isn't that a deliciously large pile of worms there? You guys need maybe one or 10 of those replacement for the current broadcasters.
- Jon Krohn: I've heard Newsmax has really reliable reporting. So I'm going to check that out soon.
- Sam Hinton: Well, best of luck with that. I mean if all else fails, I know these days, you can just open up Facebook, and they're great at fact check and you can just take whatever that creepy old uncle has shared. And that's got to be true.
- Jon Krohn: Yeah. I mean, I've been on Facebook recently, and I've seen, which I assume is the same news that everyone else sees. I've seen that everyone agrees with me on all of my opinions.
- Sam Hinton: Yeah, I mean me too. It's great. I have one uncle. He disagreed with me once. But I don't follow him anymore. So now everyone does. It's a beautiful thing, the internet.

- Jon Krohn: Yeah. And I know that my opinions are reliable and true. So if everyone else is saying the same thing, yeah, I think it's a closed case. We're all good. So anyway, we digress.
- Sam Hinton: You don't want to get too far down that rabbit hole, do you?
- Jon Krohn: Yeah. So last time you were here, we talked a lot about COVID-19 data pipelines. So COVID-19 is still here. We're still going to be dealing with this for a long time. When you started working on that project, COVID was relatively new. So the episode, the last episode was released in May, so you probably recorded it around April 2020. At that time, COVID was only a couple of months old. How did you get involved with that project and what were you doing? I mean briefly, so that we're not recapping the entirety of last episode, but just so that we have some context for what's happened since.
- Sam Hinton: Sure. So this was when I was at the University of Queensland. And UQ has a rather large medical research side to it. I heard about a senior researcher that was looking for various people to help out with a COVID-related case. I had a meeting with her. She was impressed by sort of my background outside of academia. One of the big issues they had is your data governance pipelines, handling all of that things that academics are generally not that well suited for. And so I gave her my background, I was put in as the technical lead for that project. And then over the course of the next several months or the course of the entire year, we were gathering data from several hundred hospital sites around the world. This was an

observational study, so that means that hospitals would sign up. And we would sort of say, "Look, when a patient with COVID comes in, can you record these details? And then every day from then on, can you record these details so that we can try and figure out what's correlated with what and which patients that got which medication had certain outcomes so that we can hopefully, at some point, try and help the medical staff come to better decisions?"

Jon Krohn:

Did it work? Was it helpful?

Sam Hinton:

It's sort of a bit of, I think the long term impact from the project isn't actually going to be about COVID-19 treatments at all. I'm hoping that the long term benefit is setting up these sort of projects so that they can get off the ground quicker if this ever happens again. There was so much red tape you had to try and cut through, and so much bureaucracy, and even just issues with how different hospitals and countries gather data, and how that data is contained in those hospitals' internal systems. And now that we've sort of shown that this is a massive problem, and it's almost impossible to extract the data you need from those systems, hopefully adds a bit of a fire under someone's butt, so that those systems get upgraded or standardized, or at least have different methods in place so that data can be requested from them because the fact that it took us six months to get to 1,000 patients, when there were hundreds upon thousands of infections is crazy.

Jon Krohn:

I was going to ask, were there more than thousands of infections in Australia?

Sam Hinton: Oh no.

Jon Krohn: Well, I guess even more.

Sam Hinton: Yeah, so the good thing was because Australia wasn't so affected, we had more capacity to do the analysis. So initially, at the start we rattled off with the Italian doctors, because that's where most of our data was coming from. These days, most of it comes from America for obvious reasons. So whilst their clinical staff are sort of swamped just trying to keep things afloat, we can use the fact that we're not swamped to put our medical staff into the research and analysis.

Jon Krohn: Cool. I didn't anticipate that.

Sam Hinton: Yeah, it's a global collaboration.

Jon Krohn: Yeah, but you're not doing that anymore. So you've kind of, so we've talked about how you've moved away from Australia. There's been quite a few changes. So you were in Queensland, which is a state in Australia. You call them states, right?

Sam Hinton: Yes, yes, yes.

Jon Krohn: I probably dodged a bullet there. And so you're in Queensland, at the University of Brisbane doing a Ph.D. shortly before the COVID project? Sorry, University of Queensland in Brisbane.

Sam Hinton: Same same, yeah.

Jon Krohn: And you've since been awarded that Ph.D. So in July of last year, you were awarded your Ph.D. for your work at UQ. And you've moved on from the COVID work. How did you end up doing what you're doing

now? So we're going to talk about that in a lot of detail. So you're now in England, working at the Arenko Group doing MLOps. What was the journey like?

Sam Hinton: Oh, now that is a torture journey, has often, in 2020, what else would you expect? So for a brief recap, I submitted my Ph.D., started working with the COVID-19 Critical Care Consortium. During that process, they awarded the Ph.D. I'd applied for various jobs within astrophysics because astro is fun. I submitted a few things. I got-

Jon Krohn: That's where your Ph.D. is in.

Sam Hinton: Yeah. I mean, I spent a lot of time learning this stuff, might as well try and use it. So yeah, I put out two, well, I responded to two fellowship positions. One in Chicago at the Kavli Institute, the Cosmological, Kavli Institute for Cosmological Physics, KICP Center. And I was offered that one. And I was also offered a position at Lawrence Berkeley labs, the Chamberlain Fellowship, which would have been great. Either of those, they're both fantastic opportunities. At Chicago, I would be able to continue working with a group in The Dark Energy Survey that I've worked with for years. At LBL lab, I'd be working with different groups, but there was a broader range of physics and because it's a much larger research institution, and I'll also be working with Saul Perlmutter, who you may or may not know, 2011 Nobel Prize winner, but he has interest in physics.

Jon Krohn: The name's familiar, yeah.

Sam Hinton: So his interests are physics, data science, but also teaching so critical thinking, epistemology, philosophy, something I'm also really interested in. So going to the States will seem like this would be absolutely fantastic. I got the offer, and then COVID hit. So I got the offer at the end of February. Essentially, to immigrate for a few years into the United States, as a non United States citizen, is quite challenging. So once COVID hit and the cities shut down, and we couldn't travel, I couldn't get to the embassy to have my interview, essentially everything ground to a halt.

Sam Hinton: I spent pretty much a year just trying to get down to Sydney, have my interview, get over to the States. Something went wrong each time. Every time Sydney would open back up, I would book flights and an interview and then it would close back down and they would cancel everything. And then you go back into the waiting queue of months and months again, which gets longer every time it closes down because people are waiting. So at the end of it, I was like all right, this is obviously not working out for me. So let's look somewhere else. I have British ancestry. So I thought, let's look in Britain, and seemed to be working well. There were a couple of interesting data science positions. A lot of data science positions are not interesting to me. If I come across a job advert that is just some new startup company trying to do ad tech again, or squeeze an extra 2% out of a conversion rate, that doesn't appeal to me too much.

Sam Hinton: If I'm going to do work, I want it to be work that I am really passionate about. And so I came across a job



listing from Arenko, and they're in battery storage and renewable energy. I had a look at what they've been doing. I had a look at some of the people that work for them. So that's your typical LinkedIn stalking, right? And at the end of it all, I was really impressed. So I applied for the role as a sort of data science position helping to straddle the land between pure data science and software, which is a weird and confusing place to be. But they accepted it. And I flew over with my wife early December, straight into their first, or probably their second really tough lockdown. So I got into London, I got to my apartment here, bought some plants and effectively haven't been outside since, so great times.

Jon Krohn: Damn. Well, the plants look great. If you're watching the YouTube version of this podcast, you can check those out, looking very healthy. You're obviously getting a lot of time to tend to them.

Sam Hinton: That's right. And it's so nice because in Australia, plants are always like brown, because here, everything's so green. Grass is green, it's weird.

Jon Krohn: This episode is brought to you by SuperDataScience, our online membership platform for learning data science at any level. Yes, the platform is called SuperDataScience. It's the namesake of this very podcast. In the platform, you'll discover all of our 50 plus courses, which together provide over 300 hours of content, with new courses being added on average once per month. All of that and more you get as part of your membership at SuperDataScience. So don't hold off, sign up today at www.superdatascience.com.



Secure your membership and take your data science skills to the next level.

Jon Krohn: So you mentioned dark energy, I don't necessarily want to spend too much time on your Ph.D. past. There's tons of interesting things there. You were the Australian rep for the Lindau Nobel laureates. You were on the Australian version of the TV show Survivor. You were a contestant on the show.

Sam Hinton: Oh yes, that was good. So I guess in order, Lindau was great. For those that don't know this, this isn't the ceremony where they award the Nobel Prize. But every year, a bunch of Nobel recipients, so people know the laureates that have been awarded the prize, come together with a bunch of students. And the whole thing is about interdisciplinary collaboration and trying to get knowledge from one area of science into another area. So I was really lucky to go to there. And I had a great European holiday off the back of that, which was fantastic. And then in terms of holidays, yeah, midway through my Ph.D., I got a call from Channel 10. They found my profile online and said, "We want you to come on to Survivor."

Jon Krohn: They found you.

Sam Hinton: Yeah, I don't know how. But they reached out. I thought it was spam initially. But I decided I was getting close to burning out. When you're working too many hours a week and things just aren't going well, I was losing my mind at one particular point. I was trying to fit a really nasty, hierarchical model that kept fitting with really severe bias. It was driving me insane. And they were like, "Hey, you want to have no

technology for a couple months?" And I was like, "Yes, please get me out of here." And the theme is great. But the theme for that season was Champions versus Contenders. And somehow I was put on the Champions tribe. I was effectively the token nerd in a group of sports legends and war heroes. And it was extremely awkward. We got in there the first night.

Jon Krohn: What were you champion of?

Sam Hinton: Academia.

Jon Krohn: Academic champion?

Sam Hinton: Yeah, apparently I just had a big enough brain. And they were like, "Yeah, this will do. You'll make a fool out of yourself on TV." And I was like, "Yeah, all right, because I'm not actually that smart." And the rest is immortalized in various videos that people find every now and then and send to me and say, "Is that you?", and I'm like, "Yeah." But it's good. The first night, we had this big Kumbaya moment, sitting around the fire when we introduced each other. And Lydia Lassila introduces herself and she's a gold winning Winter Olympian. Various people introduce themselves, Mat Rogers, legend. The person before me was Damien Thomlinson, who was a Special Forces commando, lost his legs in Afghanistan. After doing that, he said, "You know what? That's not going to stop me." He's known for being an actor. He's a motivational speaker. He's a very good Paralympic snowboarder, absolutely crazy dude. They've all introduced themselves, then it's my turn in the fire. And what do you say following on from a warrior or like this? I was like, "Yeah. Hi, everyone. I'm Sam and I studied physics."



Jon Krohn: It's like a hierarchical model. There's a ton of bias.

Sam Hinton: That's right.

Jon Krohn: You haven't seen hierarchical models that are as biased as the ones that I built.

Sam Hinton: It's my special skill for sure. But yes, effectively they all got on really well. They're all sporting legends or like super fit. And I'm there, this thin, weedy guy just trying to keep up. I did pretty good. So far I was undefeated.

Jon Krohn: [inaudible 00:20:24] Right.

Sam Hinton: I was the only one there.

Jon Krohn: I guess you didn't win, because that probably would have already come up. So how did you get voted off the island? You were voted off the island, right?

Sam Hinton: Yeah, I got voted off. I made a very simple mistake of trusting someone, which if you've seen Survivor, you sure don't do that. So there was one contestant. And I went up to her and I was like, "Look, I've heard that your name is going to be out there. So you might want to talk to X and Y, smooth things over." And so I'm not supposed to have told her anything because these are sensitive details and I've revealed her. She's under an NDA. She just doesn't realize it. She goes up to the other people. She's like, "Hey, Sam just said that my name was out there." Under the bus I went. They're like, "Oh, Sam's leaking information. Let's get rid of him." And I was like, "Why would you tell him? If I tell you this, just keep it to yourself. Come on." Anyway,

no regrets. I mean, obviously didn't win, so regrets, but no regrets.

- Jon Krohn: Well, I guess there's a lot of competitors. The odds of winning are, I suppose slim anyway. But let's go back to your Ph.D. When you came back from Survivor, I'm sure you had everything figured out for that hierarchical model. So you mentioned that you were working on dark energy. What is dark energy? How does that relate to dark matter?
- Sam Hinton: All right, so you got to spare a couple hours. Let's do this. So the first thing is dark energy and dark matter, completely separate. And they should have picked better names for it. Because I often give a talk about dark energy. And someone says, so dark matter. I'm like, "No, no, we're not talking about that." Dark energy super simply is, whatever it is that is causing the phenomenon that the universe is expanding. But more than that, that expansion is accelerating. So it's effectively some sort of anti-gravity force pushing everything in the universe apart from each other. And we really don't know what it is. Einstein made a guess. Currently, it looks like he might be right. That's the best guess so far, which is a bit frustrating because Einstein, hasn't he guessed correctly enough? Can't he leave something for the rest of us to try and figure out? He's not even alive. But he's still coming in like, oh yeah, I guess he was right. Yeah, he wrote this doodle. Yeah, he's right as well. It's just disheartening, you know?
- Jon Krohn: Yeah. I mean he's a tough person to compare yourself to, I guess. I think he had an unfair advantage that he

didn't have computers and the Internet, which is a huge advantage. Because now we just get distracted all day by LinkedIn and making podcasts. Imagine if all you had to do all day was make doodles. You'd have such great ideas, Sam.

Sam Hinton: Oh, I mean I want to say yes. But after coming back from Survivor, where I had a month without any internet, I did have ideas. I wrote them all down. Let's just say that they weren't the best ideas. And none of them really took off into a business to rival Amazon or Google, so potentially, I'm just yeah.

Jon Krohn: Maybe decades after your death, they will finally, someone will find that doodle and displace Apple as the most valuable company in the world in 2050. Well, hopefully you're not dead by 2050. I don't know, I just picked a number that sounded really far away.

Sam Hinton: Obviously, the idea would have [inaudible 00:24:00] killed me for it, taking it. And you know what? Respect to them. If it's worth that much, just let me know. I'll give it to you.

Jon Krohn: All right, so dark energy, but you were applying machine learning to study dark energy, other astrophysics concepts? And so you have a strong computing background. So you have an undergrad in computer science, which must be quite valuable in the machine learning space and maybe even more so in academia, as you alluded to, with the COVID situation where you had a bunch of academics saying, "Hey, we've never actually had to put something in production."

- Sam Hinton: Oh, yes.
- Jon Krohn: You have, so tell me about the things if you can weave all of these ideas together, your software engineering background, your formal education in that as well as experience working in finance as a software developer, how you were able to use that to create machine learning models in the astrophysics space?
- Sam Hinton: Right, okay, okay. No pressure, a lot of topics, small amount of time, we can do this. So the big issue that we have in academia is that, as you said, most academics don't have any formal coding background. And that means that they might be great if you ask them to write a MATLAB script, because they learned that in undergrad, or potentially the more skilled ones can write a Jupyter Notebook. And that's great. But the big issue in astrophysics is that the universe is really, really big. And unfortunately, it's not exactly like you can download the universe into your notebook. So when we have things like the dark energy-
- Jon Krohn: How big is it? Isn't it like one of the basic [inaudible 00:25:38] how big is it relative to like an ocean? Is it bigger than an ocean?
- Sam Hinton: Twice as big, twice as big.
- Jon Krohn: Oh, wow.
- Sam Hinton: It's crazy.
- Jon Krohn: All right.
- Sam Hinton: Okay. Where was I? So right, universe-

- Jon Krohn: There's so much data from the universe being twice the size of an ocean.
- Sam Hinton: Yeah. Okay so we have our telescopes. And if the weather is nice, they go and observe the night sky. They look at various patches, they take pictures, or they get spectra. And at the end of it, you have terabytes or petabytes of information. And somehow, you have to try and extract value from that. So one of the things that I was looking at was supernova. So exploding stars, unfortunately, we only want a very specific sort of supernova. They're called type 1A supernova. And that's because they all pretty much look the same. We need some sort of event that is repeatable, or at least standardizable so that we can do science on it. If they were completely random, how do you extract properties from them? You can't. But unfortunately-
- Jon Krohn: I like that it's called 1A.
- Sam Hinton: Yeah. We don't want the 1Bs or the 1Cs, or the 1BCs and all the type 2s. Those are crap. Actually, that's not true. People are trying to use type 2s, just much more difficult. Anyway, type 1As, that's what I wanted. So how do you identify a supernova going off in a galaxy? Well, it's a point on the sky. That gets a bit brighter, and then it fades away. That's it. Unfortunately, there's a lot of things in this big old universe of ours that change brightness. So there are some stars that pulsate, and you have quasars. So the black holes and centers of very young galaxies accrete matter, they eat matter, accretion disk, they get bright, they fade, they dim, they grow, et cetera, there's a whole bunch of

things. And we want to try and take all of them. And then with some intelligent algorithm, throw out everything that isn't a 1A, and that's where a lot of the machine learning comes in.

Sam Hinton:

So that's for the supernova side of it. There are other things that we were looking at within The Dark Energy Survey. So if anyone wants to see some cool pictures of something called an Einstein Cross, Google it. But what that is, is strong lensing. I'm not going to explain that. We don't have the time. But effectively, there's a way that gravity's manipulating an object such that it has a very peculiar shape. We don't have enough human eyes on the planet to scan through the images we take. So again, you have to write machine learning algorithms that take the images and try to identify these specific objects. So we use the machine learning as I guess one part in a much larger model, such that the machine learning sort of goes in near the start. And then we apply various statistical techniques or hierarchical models if you get lucky, or ABCs or approximate Bayesian computation, to try and figure out exactly the properties of the universe that would give rise to the event that we've seen. It sounds simple but not actually, that [inaudible 00:28:39].

Jon Krohn:

It sounds really interesting. I assume that you're the person who discovered those Einstein Crosses. I'm curious as to why you named it that but-

Sam Hinton:

I just felt like it was the right thing to do.

Jon Krohn:

So I mean we don't need to necessarily go into a huge amount of detail on this. But when you're building these kinds of tools, when you're doing machine

learning on that terabyte scale as an academic, I'm suspecting you're not using MATLAB scripts or working in a Jupyter Notebook. What kinds of software tools do you use? Do you use cloud computing? Do you have [inaudible 00:29:20] that you do?

Sam Hinton:

Yes. This is the main difference between industry and academia, which is, and this is definitely changing. Academia is sort of catching up to industry. So a lot of our processing works on Midway, which is a supercomputer in Chicago or NERSC. So that's one of the U.S. government's really big supercomputers based at LBL. So it's not cloud computing, but it is distributed computing. So we will get allocated a certain number of like million CPU hours. And then it's similar to if people have used Spark or Dask, but other ways of breaking up a very large data set into chunks then you process each chunk individually. That's effectively how it works, and the goal for all of this is that once you've done the analysis, once you've processed this data, the output should hopefully be able to be run in that sort of cosmological fit on your laptop. So after you've identified what's the 1A and what isn't, the collection of the one you have at the end is only like 1,000 objects in size. That's a very small data frame. And then you can play around with that really quickly on your laptop. If you ever want to reanalyze what is a 1A or what isn't, slightly bigger issue.

Jon Krohn:

Interesting, it's so interesting that after that big processing, you get just this little data frame. So you'd have, you have like a Pandas data frame on your

laptop with 1,000 rows. So you've got, top row is supernova 1A1. The next row is supernova 1A2 through to supernova 1A1000? And what are the columns of data?

Sam Hinton: So this is going to get into the weeds a little bit. I'll keep this one brief. But there are various ways that people have come up with over the years to try and parameterize supernova. So some of them use decomposition, so things similar to PCA. Some of them or most of them actually try and add a little bit of physics into that so that you have like a physically motivated decomposition. So for the supernova, you might simply have what [inaudible 00:31:25] was that, how bright it is, what color it was, and how long have we known it for. And those four parameters describe about 95% of the variation within the supernova population.

Sam Hinton: The only trick is knowing how to get those four. So there's various research going on. So there's a group at LBL that I was going to join that's using auto encoders to try and go from our time series photometry data, so how bright was it at this date. And effectively, for those that have made auto encoder, you try and predict the input, but your neural network architecture looks like two funnels stuck together. So in the middle of it, there's only a few neurons. But if you train it, and it works properly, once you get to those few neurons, you can inspect them, and you've effectively done some sort of dimensional reduction in a very nonlinear way. So it's ways of doing like that. There's like Isomapping, where you have manifold surfaces. There's PCA related

ones. There's a whole bunch of different things. We sort of try them all and hope that one of them is good enough.

Jon Krohn: Oh man, that is cool. All right. So now I'm going to have what's probably a really dumb question. So you find these attributes that are really important about supernovae. Then what do you do with them? What insights can you glean?

Sam Hinton: So fundamentally, if you have a universe that has a certain amount of dark energy that acts in a certain way, and a certain amount of dark matter that acts in a slightly different way. And you put those together, you have a deterministic view.

Jon Krohn: I told you they were related.

Sam Hinton: Well, they're different. They're ingredients that go into the pie. So there's a deterministic relationship that says, if you have your universe configured like this, and you're at this redshift, redshift is just a measure of distance, how far away that supernova or galaxy is, then it should be this bright. So you can think of it like if we compress this down, you effectively get a line plot, where we have our data is redshift and brightness. And if you have the universe a certain way, you get a line, and you just fit the line. Obviously, there's a few more dimensions, a bit more subtlety. But that's what it boils down to.

Jon Krohn: I understand perfectly. That sounds really interesting. So now I've got another even dumber question. So obviously, knowing about dark energy and how the universe is expanding or why it's expanding, it's

interesting because it's interesting in and of itself. I agree 100%. Are there also potentially applications that you know about as a result of understanding this process better? For example-

Sam Hinton: Not what we know about right now.

Jon Krohn: Time travel.

Sam Hinton: Yeah, you'd hope so but not really. So I guess I've heard this question a lot, especially when I was in Australia. So what's the benefit of doing it apart from the whole knowledge for knowledge sake, but a lot of the the techniques that we developed for astro get used in other places. A big one is genetics. So we both have very high dimensional surfaces that we have to fit or optimize. So techniques that we developed, get used. And we're like, we still have their techniques as well. But effectively, the thing with astro is that we're not launching money or resources into space. We're paying people to solve hard problems. And this harkens back to the Lindau meeting as well, which is that there are a lot of similarity between the hard problems in one field and another. So a lot of the value from astro comes from that transfer of knowledge. Sometimes you get benefits out of it like digital cameras, we got that one for free. WiFi, almost, it's a bit related to to astro, the whole sort of wavelengths and everything. So sometimes there's really cool technology that comes out. Most of it's algorithms.

Jon Krohn: Wow, man that is cool. I'm really glad I asked that question that I thought was going to be really dumb, but I learned a lot from the answer. Okay, so that was your Ph.D. And it's now been about a year since you've

wrapped up working on it. It's been over six months since you were formally awarded the Ph.D. Congratulations, Dr. Hinton. And now you find yourself in London working at the Arenko group. You mentioned that they work on battery storage and renewable energy. That does sound to me personally, people who have been listening to the podcast, I often say that using data science for some kind of social benefit interests me personally. And so it will not surprise listeners that this is something that interests me and in fact, on the preceding episode, so on Episode 459, that will air before your episode airs, Sam, the one that people are listening to right now, weird time travel stuff happening right now in podcasts.

Jon Krohn: So on the episode released a week ago, from the time that you're listening to this, if you're listening to this, relatively after the episode was released, which may very well not be the case. Anyway, Episode 459 is all about using machine learning to combat climate change. So it sounds like this is on theme that Arenko Group is in that space, battery storage and renewable energy. We didn't talk about battery storage too much in Episode 459. So I'd love to learn more about that. How can machine learning make a difference in that space?

Sam Hinton: Yeah. Okay, so a whole bunch of topics in here. But yes, I'm totally in your camp too, where I wanted to use data science on my career to do something that I felt was improving the planet a little bit. And that was one of the big reasons I shifted out of astro into

something like a company like Arenko. So the thing is, here in the UK, there's quite a bit of solar, especially during the summer months, because it is bright all the time, really weird. But there's also a buttload of wind. There's more wind being generated here than I think anywhere else. Unfortunately, the wind blows, but sometimes it doesn't. And so there's this big intermittency problem, where you can't exactly turn off the gas and the coal plants, because if the wind stops, they can't just turn on instantly. So if the energy storage, or sorry, if the energy system operator thinks that there's not going to be enough energy in five hours from now, it can't wait until five hours and then check. It has two hours beforehand, tell fossil fuel plants like coal stations, gas stations, "Hey, I think I might need you in three hours. So start turning on." And once they turn on you, they also just can't turn off.

Sam Hinton:

So once you pay for a fossil fuel plant to produce energy, you can't say no. It's going to produce that energy and you have to buy it. So the idea with battery storage, is if we kind of undercut the fossil fuel plants, so we effectively buy energy overnight, so when the winds blowing, and no one has their lights on and the energy demand is low, energy is cheap. Sometimes it's actually negative pricing, so they can pay us money to take the energy. And that's great, because we then put it in these big batteries. You wait for something like the 6 p.m. peak when everyone's getting off work, and suddenly the power demand skyrockets. And instead of having to have the gas turbines spin up at like 2 p.m. you just say, "Look, we've got it. We already were

selling the energy here at 6 p.m. that we bought at 2 a.m." And so the idea is if we can continue to increase the volume of batteries in the market, we can essentially smooth out that supply and demand curve. So it's similar, if anyone's, if any electrical engineers are listening, we're just adding more and more capacitors to the system to try and iron out any wrinkles. And hopefully, that means we can turn off the coal plants permanently. And then after that, we can turn off the gas plants as well. And I realized-

Jon Krohn: So Arenko Group actually owns, yeah go ahead.

Sam Hinton: Sorry, I realized I haven't talked about the machine learning part of it yet. The machine learning comes-

Jon Krohn: Yeah, I was going to get there. Don't worry, I would remind you but before we get there, so the Arenko Group actually owns these batteries.

Sam Hinton: No. So they used to. They own the Bloxwich battery, which was sort of a giant proof of concept showing that you can have this warehouse size battery, and it can be really efficient. But they've pivoted into a software company. So Arenko sold off the battery, used that to essentially expand their data science and software team. So the idea is there are many, many companies that are investing and building their own battery solutions. But there's a lot that goes in. So you don't just have a battery, and you don't just connect it, wire to the grid, there's a whole bunch of stuff you have to do. Because there are energy markets, you need to say that in this market at this time, I'm buying this amount of power at this price. And you have to try and figure out what markets to buy in, what markets to

sell in. How do you do that? How do you optimize your asset, so that you're actually making money for yourself and undercutting everyone else? Because that's how you make the money, you make sure that you just get in and the gas plants don't.

Jon Krohn: Right. Okay, cool. All right. So now I understand about Arenko. Now tell us about the ML applications. So tell us, we can start with, I guess, machine learning generally, and how this is useful with these battery systems. However, then also please expand and tell us about the MLOps that you focus on and maybe even telling us a little bit about what MLOps is.

Sam Hinton: I can definitely do that. So back in the good old days, and this applies to energy. But also, I think more intuitively, it applies to finance as well. Imagine you're on the stock market. Back in the good old days, you have a whole team of traders, and they read the articles, they monitor the market, and they're the ones that say, "This is going to go up, we should buy this now, sell it off later." People are expensive. And having a trading team that works 24/7 means you don't have one person. You have multiple people. They might be sick, you have even more people and they all have to have a lot of experience. These aren't easy markets to predict whether you're in energy or finance.

Sam Hinton: So traditionally, you have this very large sunken cost, that isn't just a one time cost. It's every day you're paying for a whole team's salary, to try and figure out what to do with your assets. So the idea with the ML learning side is, instead of giving the data to a trader, and to be fair, we still do have a trader. Joe is a lovely

guy, and we use him to get feedback as to how well we're doing. But instead of giving the data to them, you give it to the computer, and the computer says, "All right, so I think from the data you've given me now, that in four hours, these five markets will be at these price points with this confidence." And you can use that information to try and determine via a different algorithm, whether you should buy or sell at that given price point or wait a bit later for a different expected price.

Sam Hinton: So they already have quite a few solutions in here. So Arenko is sort of leading the pack right now, broken into multiple markets. And they're doing very fancy things like trading on two markets in one, both at the same time. That allows you to do very nifty things, that I'm probably not allowed to talk about those details, but it's very exciting. However, one of the big things that you have in many companies is that you have a whole bunch of great ideas or things that you want to do. So in our case, they're new market. So maybe we want to try and expand overseas, like Australia is a great place to put solar and batteries because we get a lot of sun, and there's enough public interest for them, even if the political interest is lagging a bit behind.

Sam Hinton: So let's say we wanted to go into Australia. You now have a whole bunch of problems, which is how do you predict the Australian specific market? Because these markets aren't the same, you need to have bespoke solutions for bespoke markets. So if someone comes up and says, "Oh, I've done some exploratory analysis. I have this great Jupyter Notebook that seems to give

great predictions." The question is cool, and are you going to put that into our production system? So how do you go from a Jupyter Notebook with a proof of concept into something that needs to run with 100% uptime up in the cloud without any supervision or whatsoever? And that's effectively the big part of MLOps.

Sam Hinton: Now, within that, there's a whole bunch of smaller things. So experiment tracking, that's definitely a big one. So the big players in here, people might have seen MLflow, or Neptune or ClearML. But the idea here is that you want your machine learning models to be a little bit less haphazard and more accessible. So I want to effectively try and take some of the concepts from software engineering, and apply them to data science. So if you think about a software project, you have it on GitHub or GitLab, you have some version control system, so you always know that at this particular commit, this is exactly what the project looks like.

Sam Hinton: That is normally lacking completely for data science projects. If you have a notebook, and you tweak a hyper parameter, and you run it again, now you've got a different model. But what happened to the previous model? It's normally gone unless you saved it out. And then if you saved it out, I hope it's not just like on a file on your computer. So experiment tracking is where you have a way of automatically saving these models, putting them into the cloud. So with something like MLflow, which is one of the biggest players in here, you can have your model, whether it's a sci-kit learn model or a PyTorch model or a TensorFlow model. And

you can say it's using these parameters. And these are the results it got and here's the model itself.

Sam Hinton: And you press a button or you write a line of code, and at the end of your notebook, it uploads all of that into the cloud. So now, if you ever want to try and recover that model, maybe that turns out it was the best. And, wow, I wish I hadn't changed those hyper parameters, you can just open up MLflow. And you'll see that at this date, you pushed this model that lives here with these parameters that performs this one. And so now you have all these models tracked up in the cloud. You don't need to worry that, ah, I had a great idea. But oops, I changed the notebook.

Sam Hinton: So experiment tracking is sort of the first big part. And then the even bigger part is how do you take an experiment and turn it into a production model. So the difference there is with a production model, you'd want to have it hosted somewhere, some API set up, so that you can say, "Look, this model that I made at this time, give me predictions when I give you this input." So the idea is you don't have to have a Jupyter Notebook. You can just make a cURL request, and down come your predictions. And it needs to be in a way that that server doesn't go down.

Sam Hinton: And also, another big issue is that machine learning models, the ones that we make, they generally take data. You give them X your data frame or your NumPy array, and they produce Y on NumPy array. Great. Not very useful though, if you're an end user or a program that's trying to use that API. So you don't want to say, "Hey, here's X. What's the predicted price?" You want

to say, "Hey, as of today, what is the price that you think will happen in an hour?" And then hopefully, your model under the hood is able to take that request, turn that into your input array, ask the machine learning model what do you think is going to happen, and then turn the output from a NumPy array into something that's more well suited to an application.

Sam Hinton:

So the machine learning model has two components now where it's the transformations of the data and the data manipulation, and then the machine learning part and they get stuck together. People are probably familiar with this, I guess the closest concept that a lot of people know about is sci-kit learn pipelines. You can have distinct steps that flow into one another. And it's quite similar to that. But it also it needs to be productionized. So you need to have it hosted somewhere in a very robust way. And you need the ability to say, "Hey, I've updated this model, updated in production, or put it to staging, have it in a testing version." There's all these little concerns that you need to have but you also don't want to try and slow things down. Every hoop that you have to jump through slows down your data scientists. It slows down the speed of their iteration. So effectively, my job is to add those hoops by trying to do it so that the hoops are as invisible as possible.

Jon Krohn:

Right. Wow, that was an incredible explanation. I was genuinely riveted throughout that entire bit. I learned a ton. And you masterfully explained MLOps in general as well as MLOps for this specific renewable energy

application. So there are a million places I could still go with you from here. I'll probably actually have this podcast episode go on forever. Well, maybe not forever. But we could make the episode length twice the size of an ocean, say.

Sam Hinton:

That's big.

Jon Krohn:

That's a pretty big podcast episode. So when you are working in your day to day, what kinds of tools are you using? Do you use Mlflow or do you mostly have to build bespoke tools? What are the MLOps tools that people should be looking at?

Sam Hinton:

Right, so I looked at quite a few experiment tracking systems to start with and eventually we did go with MLflow. It's got a large community and that really does help if you run into any issues. However, that being said, MLflow might not be the way that anyone listening to this would want to go. Experiment tracking is a very, very busy area of development. Every week, there's a new experiment tracking tool that offers new exciting features. Because of that, who knows in six months, or in one month, whether anything [inaudible 00:50:21] for us MLflow, easy to install, easy to get it working. And it came with a few features that some of the others didn't. So it has registered models, which means that I can say that this particular, so you do hyper parameter training, you have 100 evaluations, so you have 100 models.

Sam Hinton:

I can say the best model of those is now going to be called this name and it's going to be staged. So you put it into dev, and it has that sort of versioning of models built in. A lot of others don't. And so we use

that, so that every now and then we have various services that live up in the cloud. And in MLflow, what are the latest models that have these states, it takes those and it automatically deploys them into the server so that we can access them via an endpoint.

Sam Hinton:

So once we have those models up there, anyone can call and get predictions. But then we also want to save out those predictions so that every half an hour, you say, "What do you think now? What do you think now?" You put that in a database. And that way, you have a really great way of saying what your historical predictions are, because you want to make sure. Models don't stay good forever. Models have a lifetime, especially in the past two years. If you look at things like energy prices, the fact that there was a lockdown, everyone started working from home, drastically changed how the energy market looked.

Sam Hinton:

So a model that was trained in 2019 does horribly bad 2020. And 2020 is different from 2021, because of Brexit coming in at the end of the year. So ideally, you want a way that you can continuously every week, just train your model again, bump it up, put it somewhere, and then you can compare it to your new model against the older model. You want to see, are the models drifting? Are they getting worse? Should we deploy this model in testing? Make it the official model in production. There's a whole bunch of little things like that, that you have to try and figure out yourself and for a lot of that, there aren't great off the box solutions, because it's so tailored to a specific problem.

- Sam Hinton: So we ended up making those services I am talking about now are all things that I've written myself. However, there are though, there are systems and frameworks that try to make this a bit easier. So Kubernetes, for example, has good flow, you can plug seldon-core in to do the model serving. That will take your model and put it on an endpoint. We don't use that, mostly because when I joined Arenko Group, they had infrastructure already set up and it wasn't Kubernetes. And I was like, "Well, I can work with the existing infrastructure and keep the ops burden on the software engineers low or I can try and request a brand new thing unlikely to get approved because of the extra cost."
- Jon Krohn: Right, right, right. Wonderful explanation. So if listeners are listening, which I guess is what listeners do, hopefully you're listening listener. So listener, if you're listening, you might have wondered, how could I be better at MLOps myself, and you might be a manager who's interested in having this in your organization as a part of your data pipelines or you might be a data scientist or a software engineer, and you think wow, I definitely need to have this kind of versioning that I'm used to with Git, with my version control with my software, I need to have that same kind of version control with my models. So is there a place that you can recommend that people go to to learn about MLOps, or do you pretty much just have to start tackling your specific problem, just kind of give it a go?

- Sam Hinton: So when I faced the same question a little while ago now, the place that I went to that was actually very handy was a Slack community. It's the MLOps.community. And they're very open, very helpful, you can join there, introduce yourself. And there's a bunch of channels where you can ask questions. There's a channel dedicated to learning resources or requesting specific help with specific problems. So I remember during the start, I was in that. I was like, "Look, I'm looking at these five tools, has anyone got direct experience with at least one of them?" I got a bunch of responses that allowed me to filter this list of like a dozen potential avenues for investigation down to around three. And then I rolled out three different prototypes, demoed those to the team. And from that, we picked one final one that we're going to run with.
- Sam Hinton: So that's definitely been the most useful place for me. I'm sure there's probably a couple of LinkedIn groups as well. I just don't use LinkedIn that much so I always have Slack open because my work uses it and it's just simply another team that I'm part of, and it's been great.
- Jon Krohn: Nice. So maybe not the biggest LinkedIn user, but you did mention open and the ideas of open source and community. So I know that that's something that's very important to you in general. We were talking a little bit before the program started, about how somebody, how anybody, how any listener could go from some software that they've written that they think the public might be interested in, and how they can

tidy that up and make it interesting or valuable to somebody in the open source community. Do you want to talk a little bit about that?

Sam Hinton:

Yeah, sure. So I have a few open source packages, mostly related to astrophysics. So not too useful for the vast majority of people listening, I suppose. But putting those out in the software engineering background, I know that there were a lot of people in astro that had these really nifty algorithms, really cool ideas. And they were just in some random Python file or notebook. And there was no idea like, how do I put this on GitHub? Or they didn't want to put on GitHub because they were embarrassed because the code wasn't up to standard. It didn't look like NumPy source code. So that therefore must be absolutely terrible. So I actually, I did a workshop on this back in Australia, in a few cities that I flew between, about how to sort of turn your code-

Jon Krohn:

You taught the workshop.

Sam Hinton:

Yeah, it was a nice little, a nice treat. Yeah, it was a good, like 100 slides. It took us a whole day. There's no YouTube videos on it, sorry, people listening, there is a repository somewhere that I don't recommend you try and find. But places like GitHub actually have their own recommendations and systems. So you can just Google up open source GitHub recommendations, and they'll show you like, here are templates for issues, templates for pull requests. There's GitHub discussions now a way that you can add that to a repository. You've mentioned that people can ask questions that aren't issues. But a lot of it is very

accessible. I'm not going to link a specific resource, you can Google that and find it out. But I guess the thing that I would say is, I would encourage people to do it, because in the act of trying to turn your idea or your script into an open source project, you will learn a lot.

Sam Hinton:

So you'll learn a lot about Git, you'll learn about CI/CD, because you probably want to hook up something that does testing whenever you push to a branch who does testing or builds images, you'll learn how to write those tests, you'll learn how to refactor your code, you'll learn how to write documentation, you'll learn how to plug Sphinx in. So Sphinx is the program that is used by everyone you know and love, sci-kit learn, SciPy, and NumPy, that turns that doc string into the web pages that I'm sure we all have too many tabs open. So it is something you can throw into your pipeline to know whenever you change the code, the documentation automatically updates. You can have examples that create plots that update, and all of these things that you learn will help you be a better, not just coder, but a software engineer. So it teaches you the principles as to why you want to do things, not just that you should be doing things in a certain way.

Jon Krohn:

Nicely said. You're preaching now, you've made me the choir. So I remember something that occurred to me. I have various open source projects of mine and I'm kind of just doing my own thing. And now that you're talking about all this, I feel like I don't know what I was thinking, that I should definitely be trying to follow some standards, and making these projects a lot

more valuable to everyone. So brilliant, thank you for that guidance. As I mentioned in my last segue between topics, I really do feel like we could go on talking about a million things forever. But we're going to start winding down the episode. And at the end of any episode, we ask the guests if they have a particular book recommendation. And in your case, I want to do something else first, which is for you to tell us about the fiction book that you are currently writing.

Sam Hinton: All right, well, no idea when or if it shall ever see the light of day. But I am a voracious reader. I got into fantasy very early. Thank you, Harry Potter. And from that, I have read, I don't know how many hundreds of books but it's one thing that I told myself I would start doing, especially during lockdown when there's essentially nothing else to do because I was getting burnt out by the constant need or pressure from myself and I feel like from the expectations from the wider community that a data scientist shouldn't just have a full time job, they should be producing content or projects or portfolio pieces off to the side.

Sam Hinton: And for me, I was like enough is enough. I spend many, many hours a week in my professional life doing data science. So let's try and not do that when I have some time off, because otherwise, I'm going to go insane. So I effectively took some of the books that were my favorite. I had to hold a lit review process in a very scientific way. And I said, what do I like, what do I not like, sort of took the things that I did, starting to mash them together, create some characters. It's a

nice world building thing that even if a book never gets published, maybe I can run a game of Dungeons and Dragons in this setting, or something like that. But it's just a really nice way of trying to change tact when you stop work, it's something that I can use to try and clear out all the code that's buzzing around in my head that I haven't yet written down so that I can actually relax, and try and enjoy the evening or the weekend, when you get some of those.

Jon Krohn: Are you writing the book in a code editor?

Sam Hinton: Yes, no I'm using VS Code, and it has a plugin called Foam, or the markdown plugin, so they're all linked to each other. It's a nice, I have my little tree diagram, which has like places and systems and characters and plot points, and then chapter one is, it's great. Anyone that doesn't use Git when they're coding is just asking to lose a manuscript.

Jon Krohn: Yeah, so my book, Deep Learning Illustrated, which is not fiction, I hope it is.

Sam Hinton: I've seen those illustrations. They're so nice.

Jon Krohn: Thank you. That's Aglae Bassen. She's a genius. So she's been a friend of mine, she is now the wife of someone who's been a friend of mine for 12, 13, 14 years, a very long time. And she is brilliant. It is mind blowing to me how I could, at that time that we were creating the book, she was living in Paris and I was in New York. And after I drafted most of the book, we would then get on phone calls, and I would describe those illustrations over the phone.

- Jon Krohn: And she would nail it almost spot on the first time. And I'm blown away. She sent me a draft of an illustration. And I was like, "How did you take that? This is exactly what I had in my mind. I can't believe that you were able to create it." And she doesn't come from a computing or data science background in any way. So I don't know, completely blew my eyes.
- Sam Hinton: Is this your way of saying that you're a good communicator, Jon? Because I hear your message.
- Jon Krohn: I don't know, I think that if anything, the thing you should be taking away is despite poor communication, it's an even bigger statement as to her capacity to think.
- Sam Hinton: That's definitely the better way to think about it.
- Jon Krohn: Yeah. Anyway, yeah. So we did that book in WayTech. And we did all of the versioning in GitHub. And I definitely do recommend it, particularly if you're, I guess it doesn't even matter, whether you're collaborating with people or not a book, having proper versioning is great. And there's all kinds of benefits to doing it in code, like being able to write comments. I don't know.
- Sam Hinton: I think VS Code does 10,000 extensions that you can plug in. There's even extensions that help you with the writing. I haven't used any of them. But I've seen them talked about in various forums, things that look at your sentence, are you varying sentences, are you using passive tone, very useful things. Because people that make VS Code and make the plugins are smart cookies, and they want to make their lives easier, and

you can steal all their hard work to make your life easier. And isn't that what life's all about?

Jon Krohn: That's what computing is all about, sure. So is it too early for you to tell us a little bit about the plot or anything about the book? Too soon? Too soon?

Sam Hinton: Yeah, definitely too soon. I have it all in a first pass. But I've told myself because I have this thing where I feel bad trashing a section. If something doesn't feel right, you're like, "Oh, let's just tear this out." It's like you put all this effort in, you have been invested into it. But you know it's probably for the best to rip this band aid off. And I don't want to add any more burden to myself by giving away anything because then that band aid sticks even harder.

Jon Krohn: I understand. That's fair enough, but I won't be surprised if that has something to do with an old time medieval physics expert stuck on the secluded island with backstabbing people you can't trust.

Sam Hinton: You're actually pretty close. So it looks like I'm going to have to change the whole plot again.

Jon Krohn: All right. And so do you actually have a book recommendation for us since we can't read your book yet?

Sam Hinton: I mean, the book that I'm most looking forward to is Will Wight's Cradle series. So book nine is coming out in like two weeks. And I'm very excited for it. The first three books are free right now, on Amazon, I believe. If anyone that wants to check it out, you don't need to pay anything. It's just a fun fantasy series that I greatly enjoy. It's super easy to read. I don't have to

worry about keeping track of 10,000 characters. It's not Game of Thrones so I really enjoy it. And I'm super keen for it to come out.

Jon Krohn: Awesome, yeah. One of my bugbears in a book is having to keep track of all the characters. The name pops up, you're like, "Who?"

Sam Hinton: Have you ever [inaudible 01:06:02]

Jon Krohn: No, I don't even quite catch the word.

Sam Hinton: Malazan. So Steven Erikson wrote a series of novels quite Malazan, Book of the Fallen, have read them all, but it is by far the most complex and the densest set of books I've ever read. They're great. But it's recommended that you go into reading them with like our reading guide. There's no exposition. You just get dumped in the middle of this massive plot. You don't know who anyone is. And the author is like, "Well, you'll figure it out eventually." And it takes a couple of books, but you get there and it's wonderful but if anyone wants a challenge with like a long burn but good payoff, give him a shot.

Jon Krohn: There you go. I don't know, Catch 22 for me, it was a book like that. And I did not enjoy the experience. And I didn't feel like it ever got to a really big point. And I think that was kind of the point of the book. Anyway, I don't want to ruin it for people. Brilliant. Thank you so much, Sam, for being on the show a third time. I think we'll need to have you on for a fourth episode very soon because I have 100 questions that stemmed from the topics that we did talk about today. I'm sure listeners would enjoy that. Thank you so much.



- Sam Hinton: Thanks for having me, Jon. Hopefully, we'll be chatting soon.
- Jon Krohn: What did I tell you? Sam is brilliant. In today's episode, we covered how terabyte scale machine learning is used to study dark energy and supernovas, giving us some understanding of how the universe is expanding at an accelerating pace, the intermittency problem of renewable energy sources like solar and wind, and how it can be overcome by battery storage and clever machine learning deployed on the energy markets. And we focused a fair bit on machine learning operations, including its value for tracking model weights and data over time, as well as allowing more reliable model uptime in production.
- Jon Krohn: We talked about Sam's favorite tools for machine learning ops today, and his recommended resources for getting started with MLOps yourself. As always, you can get all the show notes, including the transcript for this episode, the video recording, any materials mentioned on the show and the URLs for Sam's LinkedIn profile at superdatascience.com/461. That's superdatascience.com/461.
- Jon Krohn: If you enjoyed this episode, I would of course greatly appreciate it if you left a review on your favorite podcasting app or on YouTube, where we have a high fidelity smiley face filled video version of this episode. I also encourage you to follow or tag me in a post on LinkedIn or Twitter where my Twitter handle is [@jonkrohnlearns](https://twitter.com/jonkrohnlearns) to let me know your thoughts on this episode. I'd love to respond to your comments or questions in public and get a conversation going.

Show Notes: <http://www.superdatascience.com/461>

You're also welcome to add me on LinkedIn. But it might be a good idea to mention you were listening to the SuperDataScience podcast, so that I know you're not a random salesperson.

Jon Krohn:

Since this podcast is free, if you'd like a hugely helpful way to show your support for my work, then I'd be very grateful indeed if you made your way to the Data Community Content Creator Awards Nomination form. The link's in the show notes. Of course, we'd hope you could dominate this SuperDataScience podcast for category seven, the Podcast or Talk Show Category. I'd also love my name Jon Krohn nominated for category eight, the Textbook category for my book, Deep Learning Illustrated. And finally, I'd also love my name, again Jon Krohn nominated for category two, the Machine Learning and AI YouTube category for my YouTube channel, which contains tons of free videos on deep learning, linear algebra applications and machine learning libraries.

Jon Krohn:

Finally, a reminder that starting with Episode 465 in two weeks' time, we will begin releasing guest episodes on Tuesday mornings New York time. Historically we've released Wednesday evenings, but by releasing 36 hours earlier, we'll be giving you two more morning commutes in your week to enjoy the episode. I can't imagine any downsides to the change but I didn't want you to be caught off guard when it happens. All right, thanks to Ivana, Jaime, Mario and JP on the SuperDataScience team for managing and producing another great episode today. Keep on rocking it out there folks, and I'm looking forward to enjoying



another round of the SuperDataScience podcast with you very soon.