



SuperDataScience

**SDS PODCAST**  
**EPISODE 982:**  
**IN CASE YOU MISSED**  
**IT IN MARCH 2026**



- Jon Krohn: 00:00 This is episode number 982, our ICYMI in March episode. Welcome back to the SuperDataScience podcast. I'm your host, Jon Krohn. This is an in case you missed it episode that highlights the best parts of conversations we had on the show over the past month. In episode number 975, I ask author of *The Next Renaissance*, a bestselling book. Highly recommend it. I've been reading it. I love it. And the author is named Zack Kass. In this clip, he talks about how AI might help schools tailor their education to each student. Do you think that AI tools can play a positive role as well with things like personalized learning in the classroom, helping people? So it means that this situation that we've had historically, if I think back to my grandmother, she was in a school where all grades were in one room and a teacher is having to figure out how to teach all grades at once.
- 01:04 Grades one through eight, or I guess in the US they would say first grade through eighth grade. I'm Canadian, Zack.
- Zack Kass: 01:11 Where did she grow up?
- Jon Krohn: 01:13 In Ontario. In rural Ontario, no electricity, no running water. Yeah, in the 1930s, 1940s, kind of like a two hour drive out of Toronto. And yeah, so her experience of education was where a teacher has 20 kids and those kids are all different ages, five to 13. And we figured out over time how to get children for the most part in most parts of the developed world, you'll now have kids that are around the same age, around kind of a one year band progressing together. But of course, even in that scenario, all these kids are learning at slightly different rates. And so the kids that are really good at math, in the math portion, they're bored. And so it seems to me like there's this opportunity with AI to be having personalized education for everyone. And maybe the teacher kind of becomes more of offering oversight and encouragement and that kind of human element, the empathy.



- 02:21 I'll turn the floor over to you.
- Zack Kass: 02:23 I mean, we could, and actually I do. The book, if it has a lasting legacy, it will likely be its ideas for parents. And I actually, I wrote about it this morning. I wrote in a LinkedIn post that I think that I didn't realize when I was writing the book who I was writing it for. And in fact, I wrote the book originally as simply a place to think about my own arguments. It really wasn't for anyone else. And I never really considered that people would read it. I didn't really consider that anyone would read it. And this is also an aside, but it's fun. In the process of writing it, I started getting really into a bunch of other nonfiction and I was just devouring my favorite nonfiction because writing nonfiction is ... Entertaining nonfiction is quite challenging. I mean, especially if you're talking about a technical subject.
- 03:26 And so I went back and read a bunch of my favorite nonfiction writers, one of which is Hunter S. Thompson, who's just all time.
- Jon Krohn: 03:34 That's Fear and Loathing in Las Vegas is his most famous title, I think.
- Zack Kass: 03:39 It's certainly one of them. Yeah. Yeah. Fear and loathing in Las Vegas. Fear and loathing on the campaign trail is ... Anyway, there's a long list. But he wrote once. The worst part about writing a book of meaning is that someone smart will have to read it. And for all of his sort of nihilism, he was quite sensitive and didn't like the idea, didn't like critics and didn't like the idea that people would ... I mean, most great artists don't. And I became paralyzed when I realized as we were wrapping up the book that people outside of my circle would have to start reading it. And still to this day, I'm quite nervous for people to like it. But I also, in the process, didn't consider who would read it.



- 04:33 Most great writers write with someone in mind, with a reader in mind. And I wrote with ideas in mind. I was like, "I need to put these on paper and stare at them." And it became very clear to me in the last probably two weeks, the book's been out a month. We sold 37,000 copies in the first month, and in the last two weeks, it became very clear I wrote the book for parents. And I wrote the book for anyone responsible for another human right now who is trying to make sense of a world that is increasingly unfamiliar against a relentlessly dystopian backdrop of news and media, who wants to feel like they are an active participant, who wants to have evidence in defense of their hope, who wants to be able to wake up over breakfast and say to their kids, "No, you're going to be great."
- 05:32 "To say to their kids what our parents said to us, or at least what our parents would've wanted to say to us."
- 05:41 "The world is going to be a special place and you're going to make it better, and here's what I'm going to do to make it better. And here's the role we all need to play." And that's the evidence we tried to give people and the conversation we tried to start. The chapter about education is the best chapter in the book. I think it's like I'm an objective observer as much as anyone. And it was my favorite to write in part because my wife is a Waldorf teacher. Waldorf is an alternative education school, much like Montessori that's a little more alternative, quite frankly, in the ways of its thinking. And I feature in the education section three luminaries, three education luminaries, two of whom are living and are very important. One is Eva Moskowitz, who has rebuilt a charter, rebuilt public education actually. Eva Moskowitz ran for education board in New York City and on reform, and they shunned her.
- 06:37 They kicked her out. She went and started this now exceptionally famous charter school in New York that



everyone should look up and learn about called Success Academy. And Success Academy has, I think it's basically the largest ... You know it, Jon. I

Jon Krohn: 06:50 Know it, not because I know anything about it, but I have this fellowship at this really cool AI startup in New York called Lightning AI. And so what this fellowship means is that I work out of their office every day, which is pretty sweet because they're doing amazing things. And next door to the Lightning AI office on 22nd Street next to Madison Square is a Success Academy Success Academy. And so I see the sign and I see the kids coming out with their uniforms on, but I actually don't know anything about it other than there's this kind of weird connection that I see it all the time.

Zack Kass: 07:25 It is a remarkable ... So it's the largest public school system in the country I think now she runs. Now it's not technically public because it's charter or it's not defined as public. And the more you learn about Eva and the more you learn about Success Academy, you learn that she's sort of become an enemy to the traditional education state. But she has built a program that is guided on accountability principles and she's doing a much better job. So I simply argue in the book, if we only hold parents, teachers, and students more accountable and treat this as a matter of life and death, which you can make the argument education is. It's like the outcome of an individual who graduates college, those who don't, et cetera, or graduates high school, rather those who don't. It's a critical moment that we should treat with a whole lot more importance.

08:22 Then I feature Mackenzie Price and Mackenzie Price is building Alpha School, which believes that it can reinvent traditional education and take it a step further than Eva Moskowitz by building schools where the student spends two hours a day at a terminal, at a screen, doing self-guided learning, as you described, where a hundred



students in the same grade at the same moment can be learning different things at different rates, and that that terminal then feeds back to the student what they're doing well, what they're not, and their guide, what they call guides, not teachers who support their education and their parents so that everyone can be on the same page about their development. And it works empirically for some students. It has a higher attrition rate than others. I call this out. I don't want to be sycophantic. Some students opt out immediately, but they have some really high performers academically.

09:17 The rest of the five hours or six hours of the student's day is spent outside in physical activity, in game theory, in life skills, and learning this idea of grit, the characters and qualities of a great person. It's working. I mean, parents are voting with their feet right now. And Mackenzie is a friend and I'm a big supporter of what she does. And I think even if you don't believe in alpha school, you can believe that it is on the right track. And then I spotlight Rudolph Steiner. And Rudolph Steiner founded Waldorf School 150 years ago and made the argument at the time very simply, which was that we were destroying, as he put it, it was the death of the soul and spirit of the child, the industrialization of education. As we hauled kids off to the modern classroom and sat them at desks and made them learn about math so that they could eventually work in a factory, that was the destruction of the soul and spirit of the child.

10:19 Now, what he didn't appreciate, or maybe he did, but didn't really calculate, was how well we would do economically, such that we would justify the industrialization of education, which we really do, we can, that we needed to industrialize education so that we can arrive at this moment. And my argument in the book is that if we borrow from luminaries like Eva and Mackenzie and Rudolph's original writings, what we can actually inform is a future of education that hearkens somehow to



a past prior to the economic optimization of a child where children, and the purpose of childhood, I argue in the book, is to understand yourself without economic incentives, which no adult gets to do. You cannot, as a rational adult, explore yourself in an open, honest way without economic incentives, but you can as a child. You can know what it means to be a human without wondering if you need to make money, ideally for most children.

11:21 And the sanctity of that experience is so important and we've lost it in industrial education so much from a young age is about getting a good job. You need these skills in order to get a good job, but maybe that's not what we need to do anymore. And it's not because I think work goes away. We can talk about this. It's not because I think we stop working, it's because I critically think that the opportunity now to give people the tools to actually explore a world that is going to be way more explorative are abundant. And to your point, the role of the teacher is no longer to be the smartest person in the room. It was never really to be the smartest person in the room. There just had to be. It's to inspire. It's to create a safe space where kids can actually explore what it means to be human in a way that sort of only childhood can offer us at that young age.

12:11 And it could restore, I think it could restore and in many ways lead to a spiritual awakening. I mean, one of the critical arguments of the book that I have to make carefully, because I don't want to sound too woo-woo outright, is I think that the next renaissance is in many ways a spiritual awakening, where people actually start to reconsider the rat race we have been on for a long time and what it truly means to be human and why we are here. And that will start with education. That will start with redesigning the classroom to actually let the child understand why they're learning, how they learn, what it means to sort of explore versus knowing something.



- Jon Krohn: 12:59 My interview with Zack really highlights why a rigorous and tailored education is so critical for students all the way up to those who teach them. I expect we will see plenty of AI agency merging in this next renaissance For Kyunghyun Cho, such agents will soon help us to direct our attention to the developments that matter. In episode number 97, I speak to the renowned NYU New York University professor about how he thinks AI will come to explore the world in a similar way to humans. When you talk about actively mining data, I suppose this is like an agent searching the web for relevant information or in the not too distant future, perhaps a physical embodiment, like a robotic embodiment, being able to explore the physical space and maybe even interact with objects in its space.
- Kyunghyun Cho: 13:43 Absolutely. In fact, all of those included. In some sense, they are not that different from each other. Say you want to run some kind of actual experiments. You want to test whether a particular protein is going to bind to yet another protein that we want to design a drug for. Now, one way you can imagine is that, well, I'm going to try all possible proteins and then see if any one of them bind to this particular target protein. Of course, all possible proteins doesn't make any sense. There are too many of them, right? We don't even know what kind of proteins exist that are stable and can be synthesized within the cells or whatnot. So we cannot simply say that we will try all possible proteins. So then what we do, we're going to use a very smart algorithm to pick only small number of the proteins that are time to test.
- 14:30 And then based on the feedback, we're going to continue to choose the next batch of the proteins so that we can actually find a good binder way earlier than trying out all those as all possible proteins. And then this is a sample efficiency we gain, right? And it's actually the same with the search as well. In the other idea, of course, we think of a search as a completely software based



implementation or the system. At the end of the day, we use Google search, we are embodied. So we are a physical robot ourselves. And then the same thing. Intranet has so many documents. One way to figure out which documents are relevant for my question is to read each and every one of them, and then trying to decide which subset matter. But again, there are too many of them. So what we want to do is that if we want to interact with the search engine to find or figure out how to find the small subset of time, and then after a couple of the rounds, I'll be sure that I have found the relevant documents.

15:26 So these things are all the same thing. Now, of course, when it comes to physical robots as well as the physical experimentations, there are a lot of issues, not necessarily because of the AI algorithms, but because physical words are much more fragile than software word.

Jon Krohn: 15:43 Yeah. You have to be careful. You can be tricky, for example, to pick a grapes without shattering them for a robot to be able to do. Exactly. But it's something that robots are starting to figure out more and more. It's an area that I'm personally excited about a lot now co-supervising this AI robotics research at the University of Auckland, and we'll see hopefully some exciting things come out of that. Definitely out of the research in general, exciting things will come out. I know Jan LaCune is bullish on world models. He's got his advanced machine intelligence startup that he has going. And even though this is a separate effort, you two co-authored a 2015 paper last year that won the best paper award at ICML in the workshop on physically plausible world models. This paper was called Planning with Latent Dynamics Models. And we'll have that in the show notes, of course, for listeners to check out, but does that tie into the conversation that we were just having, that research?

Kyunghyun Cho: 16:38 Yeah. And then it actually ties into all the things that we have talked about a bit, like the information processing



and whatnot is that ... So what does it mean for us to know about the word? Because we talked about the word model, and then it turned out that alone is a very big question. Some people believe that if I can imagine what's going to happen in the future at a very high, let's say quality or the high fidelity, then I may be able to say that I understand about the word. Some people say that that's not true. We actually don't have a good picture of the imagination that is high fidelity. We have a very high level concepts that are extracted. We know how those concepts interact with each other, and then that's probably enough. But of course, who's to say which one is better?

17:27

But nice thing about this large scale machine learning or the AI era now is that in many cases we can test them. So we all have heard about and then maybe tested those OpenAI GPT models or the Google's Gemini models and all those video generation models. I think the runway has one. These are amazing models. And then some people who are building these models in their mind have this thought that by building this kind of video generation model or the action video generation model, maybe we can build a machine learning model or the AI system that understands the word. And then using that kind of imagination to plan out what's going to happen. But then on the other hand, some might think, and then it's very natural for us to question whether that is necessary. Say I want to travel to Paris. I'm not going to plan out every single step and then trying to imagine every step I actually take from here all the way to JFK and then walking toward the gate, walking to my seat, having a sit.

18:29

I don't really care about that. What I'm going to imagine is that they say I already made it to Paris. What am I going to do? What is going to the first restaurant that I'm going to go into and then trying to order some wine and whatnot. But I don't really need to imagine all those steps in between. So then one might say that actually what we need is a very high level picture of how the world works



and then be able to jump back and forth. And then that's what we meant by the latent dynamics. And then that's the kind of foundation on top of which what Ian has been calling as a job that is a joint embedding predictive architecture. And then also this actually idea, of course, goes back decades. Many people in neuroscience have been thinking about it and control theory has been all about, can we find the abstraction or the small number of the knobs that actually matter in controlling any of the systems?

19:22

And then we work at that level. Then we're going to project it down eventually. So which one is right way to go about? It's unclear in my view, because the nice thing about predicting every step of the way makes computation extremely regular. What that means is that it's very good with the current digital computers to implement and scale up. On the other hand, just intuitively saying we don't do this kind of a say step by step, let's say imagination. So there is a kind of hope that maybe we really don't need to do that because we have a proof of concept here is that yes, we can skip all those steps. Which one will be right? It's unclear, but something will have to be done in this latent space rather than the pixel space.

Jon Krohn:

20:05

As Kyunghyun says, it isn't enough to replicate our reality and expect our intelligence systems to operate to their full potential. Intelligence systems will require abstract states as well as raw observations for us to see meaningful development. With this in mind, let's move on to the people who are developing the AI systems. In episode 973, I ask the AI performance engineer and famed O'Reilly author, Chris Fregly, how he uses automated processes and code scripting in his work. Speaking of tips and ways that people can be getting ahead more quickly, you mentioned to me prior to us recording that you've been loving using AI coding assistance. And so tools like Cursor, CloudCode, Codex, how has your workflow



changed as a result of these kinds of tools? How does that impact, in particular, AI systems performance, optimization? And yeah, what do our listeners need to know about where this is going?

- Chris Fregly: 21:04 For sure. I tweeted about this a couple weeks ago. If you're manually writing code in this year, 2026, you are way behind. And I would not have said that if I didn't spend all of last year watching this sort of evolution, because I started off in my sort of legacy ways where I was writing every line of code and I vowed for 2026 to only use these coding assistants and to see how far I could get. Now, I've done a couple of projects for some friends and for some portfolio companies and just to make sure that ... So the short of it is, you can fire off about 10 to 15 different things, like different aspects of either features that you're trying to build or bugs that you're trying to find. I'm personally using these tools right now to do a lot of optimizations and to look at different aspects.
- 22:10 And so I'll say, take a look at the occupancy percentage for this particular kernel that I'm trying to optimize. At the same time, I'm having another GPU that has the same code that's analyzing a different part of it. And so think of a four GPU system and you can run separate experiments, each one running on a different GPU or even four separate GPUs, like separate nodes that have their own memory and stuff. But yeah, my personal workflow, I'm going to be a little controversial and say I prefer Codex. Yeah, I prefer the OpenAI. One thing, and again, very, very workload specific, I would say. I think if I was doing UI stuff, I would probably maybe prefer Claude right now. Cloud's great with the UI. It's kind of my little dirty secret that I use Codex. And so while all the other developers are flooding the anthropic GPUs and the inference stacks, I've got kind of my little group of people that just use Codex and I still have really good



performance until people realize that Codex is actually better for other stuff.

- 23:25 But for right now, I'm enjoying good performance and a fair amount of token. But the one thing I would recommend for AI systems performance would be be on the machine. Don't try to do this stuff on your MacBook and hope that it works or even on a personal like Nvidia GPU, like an RTX, I don't know, 50. I've literally never used any of the personal GPUs because to me, the profile is so different. The hardware is so different. The memory bandwidth completely different. So if you are working on an AI project and you are ultimately going to deploy on a Blackwell or on a hopper, you have to be on that machine during development. Don't try to take any shortcuts. I recently tried the DGX Spark, which is the little mini one that got a lot of buzz the end of last year. There's so many things disabled on that, that none of my benchmarking was even working because it's just completely missing a lot of core hardware components and then software components on top.
- 24:34 The driver isn't the same. It's a lot of different things.
- Jon Krohn: 24:37 Nice. Really appreciate your insights on what you're doing with the coding assistants today. When you're working on those, Chris, do you still review all of the code before it goes to production?
- Chris Fregly: 24:49 I did. Yeah. Yeah. I was once like you or once the question that you're asking. In fact, I was just working with someone this weekend trying to get them a little bit up to speed. It's a good friend of mine and I'm like, "Look, you're doing things like you're going to be outdated
- 25:12 By St. Patty's Day." Yeah, it was the joke specifically. And I was watching their workflow and these assistants, they like to write little snippets of Python code or Bash scripts to get stuff done. And this person was literally reading



every single line of the Bash script. It's happening during the thinking process and they would go up and they would expand it and they would be looking and they'd try to find the script on disk. And I'm like, "Dude, you have to let go." So the short answer is I did and it's exhausting and I've learned to step back and there's like temporary scripts that these things write to write the final code and you could do that and you could be very disciplined and like review all the lines of code, you could write tests, you can have the LLM actually write the test as well too.

26:20

I've actually even sort of let go of unit tests and this is very controversial for all of my test driven friends and folks that are listening. I focus on the evals. So I clearly set up and provide examples that I know are correct. Those are sort of the collaboration set. And then I use the LLM to actually judge the quality. I have it constantly running. So I've got evals that are always running in the background in a separate tab while I'm building the software. So short answer is no, I stopped doing that and it really slows things down and it's hard for people to do, but I'm shipping code a lot faster.

27:13

With kernel optimization, you have to be very careful. These models want to hack the reward, or it's called reward hacking, and just get the fastest thing possible. And if that means zeroing out everything to make the computations a lot faster, it's going to do it. And so one thing I spent a lot of time on with my GitHub repo is these like correctness checks and correctness verifications. And these pop up ... Yeah, every single month, someone like released some small little startup releases something that they have achieved 50X speed up. And the first thing I do is take that code and I put it into my harness and boom, I see that they're not using Kudastreams properly or that they have ... It's not their fault, it's just that they've given the LLM too much freedom. And so we have to get better about what it means to review.



- 28:12      And I don't think even going through all the code that you would be able to catch these things because I assume that they went through and looked at the code, but not unless you're actually running the code on the machine and you've got all of the profiling set up where it can show you how many streams are being used and all the different aspects. If you don't have correctness checks, if that's in the form of evals for your sort of end user application or performance metrics and like real correctness checks within your performance harness, then you're not doing it the right way.
- Jon Krohn:      28:49      At SuperDataScience, we rightly predicted the hype around AI agents and how they could help us code better. In episode 971, I speak to the CEO of Fireworks AI, Lin Qiao, about where she sees the crossover between AI agents and what she has termed autonomous intelligence.
- 29:05      So we're here to talk about Fireworks AI, your business, which has done incredibly well. I mean, you've just grown so quickly. I believe you've now raised over \$300 million in venture capital, including a recent \$250 million series C, if I got that correctly.
- Lin Qiao:      29:25      Right.
- Jon Krohn:      29:25      And so it's a platform built around open source model deployment at scale, and the Fireworks AI platform is built around open source model deployment at scale and this idea of autonomous intelligence. Tell us what that means, Lin.
- Lin Qiao:      29:42      Yeah, sure. You're right. We raised our last round last year and we are growing really fast. So our mission is autonomous intelligence. This mission is very complimentary to AGI, where Here, the direction of AGI focus on investing in directing a lot of intelligence into this one model and have this model be able to solve very



different kinds of tasks in a great way. So the idea is you just build your application on top of the AGI model as a utility. AGI is a great direction. It's very scalable if it's successful, but the reality is only a very small fraction of data goes into the foundation models for AGI. If you look at words data, majority of the data, by majority, I really mean more than 90% of the data. It's actually not in the public domain. It's not in public internet. It's not labeled by labeling companies, which goes into the foundation models.

30:58 And majority of those data are private data, locked inside applications and enterprises. And data, we all know data is intelligence. Data is knowledge. And those application specific enterprise specific data is not accessible by the AGI labs. And we just leave a lot of intelligence on the table. My prediction, and that's where we're banding on, the future is to be able to activate those private data and let the model absorb additional application specific intelligence and bring the model to the next level. And this kind of motion is more like customization. It is the model and the inference deployment. We're customized towards applications in their specific pattern. And this customization should not be just one time. Our application enterprise product, it keeps evolving. So this customization should be continuous. And ideally, this continuous customization should be fully automated. That's autonomous intelligence. We are making great progress towards that direction.

32:17 And we believe the future is not one model result. It's going to be millions of models, one per application, per use case.

Jon Krohn: 32:25 All right. So the term autonomous intelligence, it sounds kind of vaguely to me like the buzzword of 2025 in our field, probably the buzzword for 2026, which is Agentic AI, but it sounds like it's quite different from Agentic AI, this idea of autonomous intelligence.



- Lin Qiao: 32:44 It's heavily connected with Agentic AI. So think about Agent is a way to automate many of our day-to-day task. So we have been living the world that many expert intense task has been gradually automated. So we can free up our time. Eventually, some of the even professions will be redefined. So for example, I think there are interview agents or hiring agents where you give a job listing, it will source the candidates and to even do the first rounds of filtering and interview for you. And there's marketing agent you give your ICP list and will source the right company, right stakeholders, and start to drive customized outbound emails and reaches. And their customer service agent just give the human agents some really good assistance to kind of be smart. And there are so many agents to doctors and so on. So this is happening, transforming our day-to-day life.
- 33:54 But similarly, another big transformation that's happening is in my domain is software development is being disrupted. And today, a coding agent can really start to behave like a junior engineer. And I'm not kidding, this is kind of really happening. And it's actually changed our interview process. And the fundamental question we're asking ourselves, is coding interview important anymore? So coding interview in the past is going to replace by how good you are at using coding agents. So it is actually happening across our day-to-day life. Now, let's go back to this autonomous intelligence era. So without that, currently, this work of continuously adapting the model and changing and customizing inference setup is done by a very, very small set of experts. So those experts are like, they have been doing AI system for a long time. They have been researcher for a long time, accumulated their knowledge over years.
- 35:23 So only a few companies who has those strong density talent pool are able to do that. And the question is, can that part be automated? And similar to others that has been disrupted and has been reshaped. And can this part



of doing product model co-design and infusing more intelligence in the model and making the inference serving tier much faster and much more efficient, can that part be accessible by a wide range of application developers without them putting in a lot of work and carrying the burden of learning all the deep knowledges? So that's what that means. So we heard a lot about AI is going to free up a lot of human labor, right? And this wave is interesting because it will start from different angle. It will actually free up a human from the high intelligence level, not from the physical level, right? The robotics is going to disrupt the fixed level of engagement, but AI is going to free up a lot of kind of high intelligence level of the tasks and work.

36:51 So that's kind of interesting change. And we are also innovating and disrupting in that space from a baseline platform space.

Jon Krohn: 37:02 With AI being able to handle tasks that require expert level thinking, it is important for all of us to consider the future of work. In episode 979, Rohit Choudhary talks to me about the transformation of job roles due to AI. Rohit, the founder and CEO of Bay Area Firm Acceldata, frames this as a period of transition towards a decentralized world. As speaking of things moving quickly and past paradigms no longer being relevant, while reading your blog posts and watching your presentations and interviews as we were doing research for this episode, it was apparent that your vision for Acceldata transcends classic terms like IT, data platforms, infrastructure and pipelines as passive tools. It's more about the enterprises, teams, and individuals who will be active participants in an intelligent operating environment where data, AI, human judgment converge to drive decisions and outcomes. In fact, we've got a blog post that I'll put in the show notes.



38:00 It's called Convergence of Personas: How AI is reshaping data management functions. And in that you wrote that AI is eroding clear boundaries, collapsing personas into a more fluid dynamic model where functions blend and expertise shifts. How should organizations rethink career paths and incentives to ensure depth of expertise isn't lost, but we have that kind of flexibility anyway.

Rohit Choudhary: 38:27

I think there's a lot of value in what AI is bringing, and there's a lot of value that humans will bring to these AI systems. There's a period of transition right now, and the transition is from going from completely human-centric systems, designs that were built only for humans to a world in which this is going to be a collaborative world where agents and humans will have to work together. It's a reality. I think people will have to wake up to that. I think OpenClaw has been a fascinating moment for all of the AI world, and I'm actually pretty excited about seeing where this world goes from here on, because there's so much of work that these agents are autonomously going to do and accomplish for these teams. Now, if you think of what that world means, it effectively means that if you are a person who has the capability of doing critical thinking and structured language, you can accomplish a lot.

39:19

So the previous paradigm had certain restrictions that you had to be a great programmer, and then on top of it, you would need the business domain and the expertise. Today, what has happened is that the level at which you can interact with these business systems, the GPUs, and the agents which can do work for you, depends upon the clarity of your thinking and your own curiosity. And if you are a person who has a lot of determination of being very clear in what you expect the output to be, along with the expertise that you know your industry better than the others, you're more likely and most likely to produce outcomes and systems which are way better than some of your competitors, peers will produce. In that world, I think one will have to prioritize the presence of these



skills within individuals who possess both of those things, creative thinking and a lot of structure, clarity of thought.

40:16 And I think that is the only way that organizations will progress further. In terms of the value of domain, like I mentioned, there's this period of interim transition between human-centric to agentic plus human-centric world. And in that world, I think a lot of domain expertise is going to get used to train internal agents to fine-tune LLMs for LLMs or SLMs to suit the purposes of your own organization, and those skills will then be required in the future.

Jon Krohn: 40:48 Nice. You used a term there, SLM. I'm sure all of our listeners are familiar with LLM, large language model. SLM, probably most listeners are familiar with that as well, but small language model. And this is a really exciting area where you don't necessarily need to have these big ... You think about Claude Opus, you don't need to have a Claude Opus size model running for every kind of task. I've had a lot of success in doing consulting for enterprises, or in startups that I've been a part of, fine-tuning very small, say, open source LAMA models that just have a few billion parameters. They can become very, very good at a narrow task when they have a high quality training data. You wouldn't ask them to do anything, but for that one task or this relatively narrow set of tasks that they're fine-tuned for, they can excel.

Rohit Choudhary: 41:35 100%. I think SLMs and LLMs will be part of the enterprise stack. I think as will CPUs and GPUs and ASIC, it's got to be an XPU architecture with SLMs and LLMs. And I think a lot of IP will start getting embedded in the SLMs as well, obviously because of privacy concerns and everything else. And if you were to just think and just abstract it out, I think AI is breaking the centralization model. When I think about this whole world, for the last 10, 12 years, the whole thing was about centrality of let's get all your data together into a



data lake, into a warehouse, into a location, into a cloud on- premise environment. Whatever's your preference? I think there isn't enough time for such large scale migrations to take place. I think AI models will have to get closer to the data, where it resides and operate on that to provide the enterprise outcome that people need.

42:33 I don't think the speed of AI matches the physical realities of large-scale migration and centrality, so it is going to be a decentralized world in the future. And if you just extend or extrapolate that argument, it's hard for me to see that only LLMs will win. I think LLMs plus SLMs will win.

Jon Krohn: 42:52 All right, that's it for today's ICYMI episode, to be sure not to miss any of our exciting upcoming episodes. Subscribe to this podcast if you haven't already. But most importantly, I hope you'll just keep on listening. Until next time, keep on rocking it out there, and I'm looking forward to enjoying another round of the SuperDataScience Podcast with you very soon.