



SuperDataScience

**SDS PODCAST**  
**EPISODE 1006:**  
**IN CASE YOU MISSED**  
**IT IN JUNE 2026**



- Jon Krohn: 00:00 This is episode number 1006, our ICYMI in June episode. Welcome back to the SuperDataScience podcast. I'm your host, Jon Krohn. This is an ICYMI episode that highlights the best parts of conversations we had on the show over the past month. My first clip is from episode number 999, where I sit down in San Francisco with the revered Chip Huyen, a two-time mega bestselling O'Reilly author. Great guest for episode 999. In this episode, we confront head on the question that so many people in our industry are quietly asking themselves. If the cost of building software is going to zero, where does the moat go? You were kind of talking about this earlier, this kind of existential crisis, which I feel as well. It ranges from across everything I do. I did a PhD in AI and that gave me a real moat around my career.
- 00:55 Other people couldn't create a machine learning classifier or understand problems with labeling data or these kinds of things. But now all of those
- Chip Huyen: 01:04 Kinds
- Jon Krohn: 01:05 Of things, a machine can do no problem. For the podcasting too, I mean, it lowers the barrier to entry. You don't have to be a great writer to come up with great topics, to script episodes. So there's a lot of people in a lot of industries who would feel like the moat is going away from them. I'm wondering, so obviously focusing on physical systems is one way to create a bit of a moat for yourself because hardware, R&D cycles are going to be longer than software and there's lots of different ways that hardware can specialize and anthropic or open AI aren't going to next month just all of a sudden have a robot that does that too. So there's a mode in physical AI. So maybe that is just the answer, but are there any other ways? Yeah. What other ways do you recommend? I guess you also had the systems design idea right at the beginning of this episode.



- 01:55      What other tips do you have for our listeners on how they can try to future proof themselves a little bit in this time? Well, nothing.
- Chip Huyen:      02:07      This is really funny because I have a friend who is an economist. He's one of the smartest people I know and he does consult a lot of governments and coming up with technical policy, tech policies on how to get the nations stay up to date in the AI era. And he was straight up telling some of them, it's like, "Yeah, there's nothing you can do. You don't have enough budget for it. Just don't do anything." Some people do have a very, very pessimistic views of the world, but I think I'm more on the automatic side.
- 02:44      AI can solve a lot of problems, but I also think that will never stop being problems for me, for us to solve. So for one thing, it doesn't matter how many AI models are there or how good AI models are. I would never stop being angry at people on the internet. There are always people that piss me off. They won't always be customer services I'm unhappy with. There will always be things just like collaboration is just not quite straightforward. So recently there's a founder and I really like the founder. He's very smart. So he came to me and he pitched this idea of another Asian orchestration framework. And then he told me that all the problems that he have seen with a lot of companies is that there's not enough communications between product and engineering, which is very classic. And he was just like, "Okay, and my agent orchestration platform is going to solve that.
- 03:42      " And I was just like, "I don't think that's a technical problem." I think easily when product and engineering people don't talk with each other, that require people solutions. It usually you don't solve that by like, okay, here's another tool. We magically make product people and engineer people get along. So there are a lot of people problems that is not quite like sold. I think that there's



several categories of tasks that I think that are not entirely completely clear follows to like how to solve. One is like human AI collaborations, right? So a lot of AI tools nowadays built upon legacy systems and things about like how we interact with like old software systems. So just like example of like coding tools. Originally we have a lot of coding tools that like just part of VS code because VS code existed. And then we have like a lot of coding tools as part of the terminal because terminal has always existed.

04:45 But then like as long as it made me think like, wait a second, why are like editing like note taking apps like VS codes and terminals? Why do you need both of them?

05:00 I mean, just going back, why is it different stuff? And another thing is like, why is terminal so hard to use? So a lot of engineers for me, like I have used terminals in school and stuff and for work. I used to, but I'm not crazy happy with it. Right. You've never been

Jon Krohn: 05:17 A big VIM person.

Chip Huyen: 05:18 No. Okay. So I have friends who are crazy VIN person. I have friends who just don't use VS code or like PyCharm or anything. They just like straight up code in the terminal. So you think find a lot faster with all the key. So in here you could use terminals as a- As an IED? Yeah, as an ID. So we know that terminals exist, but like because of like coding tools like cloud code, a lot of product people or like people who never use terminal before are suddenly exposed to terminals and they was like, "Okay, why is this so hard to use?" It's very painful. And I think it's like, okay, terminals, maybe that terminals are hard to use on purpose because terminal is actually very powerful. Terminal basically give you access or control plane for you to like control the computer. You could easily remove RF, like everything, right?



06:08 Yeah, you can just do that. So maybe like you make it hard to use so that only people are willing to get used to it use it so they're less likely to make mistakes. So I think maybe, but also like a legacy thing and I think like why don't we have something in between like we have something that can be both very, like can give you access to the far system assesses a computer, like a control plan terminal, but also easy to use

06:38 Like an IDE. And I think like that is, I talk about it actually like OpenAI and actually I introduce a bunch of like desktop app, right? The College desktop app, which is basically the same idea of like, okay, very easy to use interface, but give you like access to a lot of things, the ways that the terminal can. So I think it's evolving and also like it's a bunch of like how to access your agent when the computer is not working. You probably have seen people complaining about like, okay, like why I have to keep my computer open all the time because my coding sessions, like my cloud coding codex are doing their things. So I usually just get into Uber and it's just like, can my computer open and look like freaking nerd. But it made me think like there's no reasons my computer should be open because it's totally run as a cloud and then we need like something that can access through the phone.

07:31 I think a bunch of people are building it like, okay, now you want the things on the phone, I need some sandbox because like how do you share context between the phone and the computer? So anyway, I'm going to like very much use like genres like she's talking with you.

Jon Krohn: 07:49 Well, I'm just wondering what, this has all been very interesting,

Chip Huyen: 07:52 But



- Jon Krohn: 07:52 What is this all... Are you saying that there's lots of, lots of problems that we can still
- Chip Huyen: 07:56 Solve? Yes. So I think it's like I use it not a great example obviously, but I'm saying it's one thing is like we don't quite have a good understanding like what is the optimal way for humans to use AI. So human AI interface is on big things, right? Another category problem is just like, she's not done. It's like just the first, it's one of my first in the 20 points. No, this is really important. This is good. Okay. Yeah. Another thing, if you will let me say it is how AI interact with the world. So we have a lot of techniques to make AI good at using tools, right? But I think like for AI to interact well with the wall, we do not just want to improve AI. We can also make the world more AI ready for websites, for apps we can make it like better documentations, better APIs that agents can call, better security, like making, okay, this kind of like actions are dangerous.
- 08:57 So maybe you should like have less permission to AI and stuff like that. But how about physical wall? So I'm not sure you've seen this very cute video of like a food delivering robot. A football
- Jon Krohn: 09:10 Robot?
- Chip Huyen: 09:11 Food delivering robot.
- Jon Krohn: 09:12 Food delivering,
- Chip Huyen: 09:13 Right, right, right. Yeah. So it's very tiny robot.
- Jon Krohn: 09:15 Not a foot delivering robot. That would be
- Chip Huyen: 09:17 Weird. Do you have HR feet to like deliver? I'll
- Jon Krohn: 09:20 Take six feet please.



- Chip Huyen: 09:26 So the robots are very cute and then the robot just couldn't cross the street. So the robot had to ask a pedestrian like, "Hey, can you press a button for me so that it turn green so that I can cross?" And the pedestrian was like, "What the heck is going on? " So I actually talked to someone who worked at one of those robot food delivery companies and he told me like the hardest part is just like how to get the robot interact with the world. So actually some cities have this like streetlight API so that the robot could connect to streetlight API so it can press the button via the API. Oh, because it
- Jon Krohn: 10:05 Can't press the button to
- Chip Huyen: 10:06 Say that
- Jon Krohn: 10:07 I want to walk.
- Chip Huyen: 10:08 Yeah. So I was saying that part of like how you could provide toolings to make the wall easier for AI to operate in. Sure. Yeah.
- Jon Krohn: 10:18 From the broad question of where moats live, we zoom into one company that has been building inside this changing landscape for years and to great effect. In episode 997, I chatted with Andrey Kurenkov, co-host of my favorite podcast last week in AI and the founding AI lead at Astrocade, a wildly successful platform that allows millions of people to create video games without writing any code. Andre walks me through how his team was effectively doing vibe coding before Andre Karpathy had even coined the term, how his whole company adopted Claudecode a year ago and the most important lesson he's learned from this. Your title has shifted over the past couple of years and so there's two journeys actually that maybe we can kind of cover in one chronological sweep. So I think that you starting, your company rather starting to work on Astrocade and doing this kind of like vibe coding for video games, it predates certainly lovable being



popular or these other kinds of vibe coding platforms being popular.

11:19 So you guys were working on this vibe coding platform before it was popular and then you within that ecosystem, your title has shifted from ML Scientist when we recorded the last episode to founding AI lead. So yeah, how has the company transformed? How has your role transformed as the company has grown and gone from just developing behind the scenes to now having tens of millions of active users?

Andrey Kurenkov: 11:44

Yeah, it's quite the story. I joined in April of 2023 just after finishing my PhD at Stanford where I was actually doing machine learning and robotics. And initially that ML scientist label made sense. My background was in machine learning and I was going to be working on the AI side of things, but we sort of shifted that label a little bit because ultimately it's not machine learning to build agents or to do prompting. Machine learning, you need to understand machine learning and what is involved, but a lot of it is understanding bigger systems of prompting context engineering and just generally even more hands-on practical problems of building something that works rather than being scientific of doing research. I think my title shifted just to reflect the nature of the work itself and yeah, in that few years I joined April of 2023, the company as a whole started on this direction of building basically what we have now, which is a user generated content platform for games back in February of 2023, this is like two months after ChatGPT came out, I think two, two and a half.

13:06 So at the time we had LLMs, we had LMAPIs, but they were still stupid relative to today, right? They're not anywhere where we used to be. So the idea of straight up code generation, one of my very early things at the company was actually experimenting with code generation and they could already write little small



functions and so on. And I remember even back in 2023 when ChatGPT was coming out, people were demoing, oh, like it wrote Pong and it was mind breaking or some website. But as anyone who's followed AI over years can probably tell relative to today, going back to 2023, LLMs were much more limited in many ways of hallucination, reliability, general intelligence, and certainly being able to write code that actually works and has no bugs.

Jon Krohn: 13:59

Sure. I mean, it's really, I would say, and I've talked about this on this show a lot and with guests and I'm sure it's the kind of thing you've been talking about with Jeremy a lot on the last week in AI podcast, but it's really since February that we have reliable text to code generation with the Opus 4.6 release embedded in the CloudCode environment.

Andrey Kurenkov: 14:21

I mean, I think I'll be a little more generous. I think people have come to realize and wake up to it more since February. And it's one of the interesting things where I remember our company, we adopted CloudCode, basically everyone starting in about June of 2025. And I still remember CloudCode came out around, I want to say February of 2025. So the realization that these LMs were now able to do coding has been sort of brewing. I think Andre Kapafi coined the term vibe coding in early-ish 2025. So it actually just hits the one year mark since the term itself has been coined. And what really happened was that this entire type of product and experience matured, right? CloudCode came out very quickly, the people kind of data in their minds doing the work with Cursor realized, whoa, this is actually on another level relative to just Smart Auto Complete.

15:26

And I remember seeing the hype and not being sure if it's actually that different. Then I tried it and a few days after trying it, I was telling everyone in the company to start using cloud code. But as you say, I think this year since February, January, as the new models came out, it's just



gotten better and better and more and more reliable. And I'm sure at some point you've talked about the Meta time horizon eval, whatever where -

- Jon Krohn: 15:59 Oh yeah, the meter thing. Yeah. I mean, I talk about it every single talk that I give, I have a meter chart in the first few slides and actually I was just recording the episode that'll actually come out next week with Chipoiian, episode 999 and I talk about the meter charts in that with her. Yeah, I'll have a link to meter in the show notes, but basically it's just showing this crazy exponential increase where like with Methos it broke the
- Andrey Kurenkov: 16:33 Meter...
- Jon Krohn: 16:34 Yeah.
- Andrey Kurenkov: 16:34 You can no longer evaluate how long the task that AI can reliably do because the tasks are too long and it's hard to evaluate. So yeah, it's been a progression where there was an inflection point about a year ago where it got to a point where they could do tasks if you were babysitting them and they could write code and it now is getting to a point where they can do it without you babysitting them and that is another sort of shift that is on top of vibe coding as a thing.
- Jon Krohn: 17:07 Yeah. It's wild. What can you tell us? I mean, you kind of just disclosed one thing there that's happening behind the scenes at Astrocade, the cloud code usage, obviously without divulging anything that would be an issue publicly, what else can you tell us about what it's like behind the scenes in terms of the tech stack building a platform that is generating games being used by millions of people?
- Andrey Kurenkov: 17:31 Yeah. I'm not entirely sure how much my CEO and CTO want me to say, but - You
- Jon Krohn: 17:36 Don't need to say much of



Andrey Kurenkov: 17:37

That. I will say, I think the truth is there's no secret sauce fundamentally, right? There's an agent that has some tools. We use an LLM. It's all the standard ingredients. So the ingredients aren't special, but the way you mix them, the way you put this whole thing together is kind of a tricky part. So you need to... We have a lot of effort to benchmark and to evaluate the way we do our harness, the way we do our prompts. And it's actually a very challenging problem to benchmark because it's one thing to benchmark like a multiple question answer test where you know the answers. When your task is like implement this Tetris crossed with a merge game that also has like a puzzle component, there's like a million possible answers. You can't even do a LMS judge because we have aluminum. At least

Jon Krohn: 18:35

One million possible ways of

Andrey Kurenkov: 18:36

Doing that. At least. Yeah. So the key to what we do behind the scenes is the finer points of how you put together VLM, the prompt, the tools and make it all function.

Jon Krohn: 18:52

Yeah. I can't even imagine how tricky that would be. This is the tricky thing I think that still provides a moat for product designers across the board, which is that when you get into any particular niche like this, there's tons of opinionated bets that you make as a development team, as a product management team and some of those are going to be wrong and you get that through having good user feedback metrics, you can kind of learn, okay, going in that direction was the wrong way, let's try this other way. And then over time, over many years, you accumulate, okay, we've kind of gone in the right direction overall. And that gives you a moat. Yeah.

Andrey Kurenkov: 19:33

Yeah. I think you learn a lot and we have learned a lot over the years. One thing we've learned over time and I'm still a blog post just detailing all the many things we've



learned in doing this over a few years, but one of the things you learn is you have to be very careful around scaffolding around AI in the sense of as we were starting out in 2023 and even in 2024, you couldn't do vibe coding yet. So we had to come up with a way to let people make games where the AI was helped out by some sort of structure, which we can call scaffolding. And so it took these preexisting pieces and it put them together and made things you could use in the game. But then you hit 2025, you get to a point where you could do vibe coding. All that scaffolding now is limiting you as opposed to becoming more powerful.

20:24

So one thing that we are very mindful of is building in a way that it is very future compatible. You want to build your system in such a way that when VLMs get better, two months from now, three months from now, whatever you built isn't outdated. And it's one of these tricky things where I think probably in startups and research and everything, you learn over time that you have to know what to keep things simple and really deeply understand how to leverage technology and build something in a way that's compatible. So I don't know if that is actually interesting, but it's something I've had to learn the hard way.

Jon Krohn:

21:08

For all this software abstraction enabling our AI systems, there is a very physical reality behind it, the data centers themselves. In episode number 1003, Lightning AI's Vice President of Infrastructure, Frank Basso, takes us inside AI data centers, AI data centers, and explains the industrial reality of how the AI sausage actually gets made. A lot of us can probably picture a photo that we've seen online of these long hallways of server racks, but is there anything else that's kind of interesting, maybe particularly interesting about an AI data center when you're physically standing there?



- Frank Basso: 21:45 When you're physically in there, one of the differentiators from a traditional data center is the noise level. These systems are very noisy. We call them screaming banshees.
- Jon Krohn: 21:57 Oh my God, I had no idea.
- Frank Basso: 21:59 Yeah. So especially within air cooled and inside the data hall, the levels range from way beyond what you'd hear at a rock concert if you're in the front row. And so hearing protection for our staff, require two types of hearing protection at all times. You have both the buds that go in your ear, you can use molded ones or not listen to your music or whatever. So something occlusional in your inner ear and then cans, right and cans being all passive, you cannot wear or use active noise canceling systems within a data hall. This is a big thing that people are like, "Yeah, put my noise canceling on. It's great." Well, if the data hall is 105 to 110 decibels, to cancel the noise, noise canceling generates 105 to 110 decibels. So that does just as much damage, you're not hearing it, but it's damaging your drum.
- Jon Krohn: 22:59 Wow. I had no idea about that. It makes so much sense now that you say it, but I had no idea that with my noise canceling headphones, I'm here thinking, I'm wearing noise canceling headphones right now. Obviously, it's not canceling 100 decibels of noise. I don't have screaming banshees in my recording studio, believe it or not, but actually that's an interesting take-home tip. For anybody going to a rock concert or whatever, you need to have passive noise, not canceling, but just suppression.
- Frank Basso: 23:27 Yeah, occlusional or suppressive. And so really it blocks the two different kinds of hearing protection, the inner ear hearing protection and then the outer ear blocks different frequencies of noise as well. So the frequencies of noise that cause your hearing damage, the higher frequency noise from the fans and the motors and the



power supplies that are humming that you can't really hear to your native ear, they're present. And so you need to block all those out for safety. We take that very seriously at all of our locations and we actually have OSHA sound studies done and we maintain OSHA compliance and everyone has to get trained to be in the data center. Even visitors, we warn visitors when they're coming like, "Hey, this is a very loud environment." And what's funny is that the occlusion blocks out so many things, but if I talk loud enough, like that loud grandma talking at you because she can't hear anymore to someone in the data center with a hearing protection on, the frequency of my voice comes clearly through, but you don't hear any of the high frequency noises or things that can damage your hearing.

24:32

So you might think that, how does anyone work with somebody else? Well, there's a couple ways. One, we have some systems that are like intercom based like racing radio style that you can talk to each other with, and we also have hand signals that we use in the data center and things like that. There's a bunch of different combinations depending on which location we're at, how loud it is. And you think, oh, the liquid cooled ones don't have as many fans. They're just as loud. They have rear door heat exchangers. They have other cooling systems and pumps and things running in the room. It is a very loud industrial environment. The liquid to chip data centers are more industrial, if that makes sense. Nutritional data centers that are pretty and they're really nice, they raise floors and they're super orderly and things like that. You look at some of the pictures you see online on liquid to chip centers and there's hoses and pipes and cables and everywhere all of ours that currently are fed from above.

25:31

So there's 20 inch water mains running through the room that are insulated so they don't make water or sweat because the temperature differential with the room.



There's hoses to every machine. It's just if you look up, you'd be like, "Oh my gosh, what is all this stuff in here?" Well, that's how the sausage is made. It's very industrial. It's like you're in the reactor room on a submarine or something.

- Jon Krohn: 25:55 It's pretty cool. Do you think that there is a higher rate of sign language fluency among data center workers relative to the general population?
- Frank Basso: 26:07 That'd be an interesting question. I don't know. I don't know, but you think there should be at some point. That's actually a really good idea even though I'm sure they use that home version of sign language if you know what I mean. Right,
- Jon Krohn: 26:20 Right, right. Well, yeah, that's interesting. I guess it's potentially a good career choice for people with a hearing problem.
- Frank Basso: 26:26 Oh yeah. There you go.
- Jon Krohn: 26:31 And so for people, Frank used the word OSHA, which people in the US probably will, everyone will know what that means, but if you're outside the US, it means occupational safety and health administration as a federal body that is keeping workers safe in all kinds of industries. Quick question for you. With these data centers being so large, do you just always get around on foot or do some people use pedals or motorized vehicles ever?
- Frank Basso: 26:56 It depends on how good your insurance is. A lot of data centers, some of these now have like scooters and bicycles just sitting all over the place. I'd say don't wear Heli's because you need to wear actually proper work boots in these locations because things are heavy and if something were to drop in your foot, that would be bad. So you need to wear a proper work attire. Do you



- Jon Krohn: 27:22 Wear hard hats?
- Frank Basso: 27:24 During construction phase, they wear personal protective equipment. So hard hats and vests and ceramic towed boots and non-flammable things during constructability and provisioning. And then once it's online, the data center technicians aren't required to wear that, but they're for hearing protection or if they're working in a cabinet safety glasses and then of course they need proper ESD protection. We issue ESD shirts for our teams. So they're wearing a shirt that's not polyester that won't spark every time they touch a cabinet kind of standard issued uniform stuff that we've been working through.
- Jon Krohn: 28:06 So ESD is like electrosensitivity something?
- Frank Basso: 28:11 Yes. Electrostatic discharge. Electrostatic discharge. Yeah. And you have to wear a wristband if you ever touch or open a box. So you put it on and then the cabinets literally have like these little light and bolt and plugs on either side, every single cabinet, front and rear, and you plug yourself into those. So you're now connected to the cabinet because with all the air flowing through the systems, they can generate static electricity. So it's for safety of the worker and for safety of the gear.
- Jon Krohn: 28:39 After the industrial side of AI, let's bring the conversation back to one of the most distinctly human skills that no machine can outsource for you, self-awareness. In episode number 1005, Gilbert Eijkelenboom shares research that I found quite surprising that only 15% of people are technically self-aware. We talk about what self-awareness actually means. We define it. We talk about why it's foundational to effective communication for anyone working with data and the simple practices Gilbert credits for building his own self-awareness. A statistic from your book that blew my mind, but I'd also love to just have you explain to me what the definition



means a bit more is you cite research showing that only 15% of people are self-aware. So what does that mean? What's the definition of being self-aware? Because I'm sure it's not something that's like binary. It makes it kind of seem easy like, okay, 15% of people are self-aware, 85% are not.

29:37

I'm sure it's more of a gradient than there's degrees of self-awareness. So what does it mean to be self-aware and presumably if that statistic is true and if listeners to this podcast are a representative sample of the population, then 85% of listeners of the vast majority right now are not technically self-aware by whatever that definition means. So I'm probably not self-aware, it turns out as well. So what does it mean to be self-aware in that definition and how can you foster more of it?

Gilbert Eijkele...: 30:10

Yeah. So I believe it has been like it was a few years since I wrote the book, but I need to think about the exact research they used, but I believe their definition was that the perception of yourself, your behavior and your thoughts is overlapping a lot with how other people see you. So if I see myself as a very confident presenter or maybe the opposite. So if I see myself as a very shy person, but other people tell me, "Hey, you talk to everyone, you have your opinions ready, you don't wait to talk, all those things, then it's not really matching. So then I might not be so self-aware. But if I'm more self-aware, I understand my own behavior. I can have that meta conversation, not just being in the conversation, but also saying like, okay, here I'm rambling a bit, or let me pause there because I want to make it more concise." So in the moment, being able to take the helicopter view and see your own behavior and action.

31:24

And it's so important because everyone is different, right? Everyone is different. And the more I learn about myself and the more you learn about yourself, the more effective



you can be in communication and relationships with other people, even your personal life. I mean, for people that have a spouse or partner, if you don't know yourself really well, it's getting difficult because they see certain behavior more accurately sometimes than others. Then you also ask, "How can you become more self-aware?" In my experience, the best ways to become more self-aware is one, journaling because it forces you to articulate your thoughts and feelings in the moment and you can look back at it. I have my journal from years ago and I see my journaling from 2020 and what I struggled with back then. And then I look at that and I'm like, "Hey, okay, I think I've grown a little bit because the things I was worried about back then don't concern me as much." So journaling is massive.

- 32:29      And then second is meditation helped me a lot because it forces me to stop. There's a lot of thoughts in my head always. I'm a big overthinker and thinking is useful in some situations, but often it's also not useful. If we're having this conversation and I'm thinking like, "How do I come across? Will the audience like this? Is he judging me? And what do I have for dinner tonight?" It's not useful, right? Yeah. I'm getting out of the moment. I'm not having this conversation anymore. So meditation and journaling helped me a lot. Sports, exercise, physical, because it gets me into my body more instead of only thinking up by my head. And you can see it as well on the street, people having their phones being fully disconnected from others. If you're in the queue of the supermarket, everyone pulls out their phone, right? Or if you go to the bathroom, pull out your phone.
- 33:33      I do it as well, but I try not to do it as often. And if I don't do that, it creates this space to let the thoughts come and also let them disappear and it creates so much clarity and self-awareness.



- Jon Krohn: 33:46 We're rounding up a great month with episode 1001 in which the tables get turned. Super data science founder and the original host of this podcast, Kirill Eremenko interviews me. Kirill asks one of the big philosophical questions I get asked all the time. "Does AGI, artificial general intelligence, require something like consciousness or can we get there just by predicting the next token? I share with Kiro what my PhD in neuroscience taught me about how artificial neurons compare to biological ones and why I think we still have years to go on the breadth dimension of AGI.
- Kirill Eremenko: 34:18 Is AGI measured solely by outcomes? Does it really matter whether an agent has the ability to imagine, think and ask humans, or can AGI be achieved just by predicting the next token? Is there something specific in the brain that you found during a PhD that cannot be recreated with the way we're going about AI in this day and age? Well,
- Jon Krohn: 34:46 There's a lot of different ways that that question could be answered. The first thing that I'd like to start with is that in order to be talking about AGI, we need to have a definition of what that is. I think the best definition, we've talked about this in recent episodes. So with the Andre Karenkov episode that came out as 997 recently, we talked about a paper on the five levels of AGI that came out from Google DeepMind two years ago and I did a whole episode on it, one of the short Friday episodes, just me solo. It's about a 10 minute long episode, episode 748. And so if people want what I think is the best definition of AGI, check out that episode. And basically what they do in that episode is they talk about five specific levels of AGI based on the percentage of people on the planet that on that particular capability, the AI system outperforms humans.
- 35:49 So it's like fifth tier AGI if it outperforms 100% of humans. I can't remember exactly all the tiers now off the



top of my head, but it's something like it's like tier one AGI if it outperforms 50% of humans. So it's nice to kind of have that concrete way of defining these tiers of AGI. But then the other key thing that it talks about, so that's kind of like the depth of AGI capability, but breadth is also important. So in recent years, since the ChatGPT moment, we've seen AI models be able to match or surpass a large number of humans on office work kinds of tasks. So coding on writing, but we're still a ways off on being able to handle lots of real world scenarios. And so there are companies that are raising huge amounts of money right now to work on world models.

- 36:46 So Yan Laka has done this with AMI. They recently did a billion dollar seed round. I think it's a record and there are other businesses. Fay Fei Lee has one. There's one in London. I forget who the founder of that was. And all of those have done massive, massive seed round fundraises because there's so much potential. There's still so much that we have to do in order to be able to have AI systems to be able to act in the real world like humans do. So we're already at a point where we're at a level of some level of AGI, whether it's level one, two, three, four, five, on the vast majority of tasks that you do where you're sitting at a computer screen, kind of isolated from other people, but where you have to have some kind of physical embodiment of AI that's exploring the world, we still have a long way to go, years to go at least.
- 37:47 And so thinking about that kind of breadth and if you want to define, like if people think about lots of people, they defy AGI simply as an algorithm being able to do everything that a human can do and that must mean more than what we can do sitting at a computer, right?
- 38:05 Even by that definition. So I think we have a long way to go in terms of breadth, especially in terms of exploring the real world.



- Kirill Eremenko: 38:12 Got it. Interesting. So there isn't something we're trying to replicate that's in the brain or like a criteria that's like, oh, there's this thing in the brain. It doesn't have to be exactly like the brain. As long as we can achieve the same outcomes, then we can tick the box of AGI.
- Jon Krohn: 38:25 Yeah. It doesn't seem obvious to me that there's a mechanism by which a machine is conscious, say like you were kind of -
- Kirill Eremenko: 38:41 Yeah, conscious.
- Jon Krohn: 38:43 Yeah. Yeah. It's not obvious to me, like I could be proved wrong in the future with some level of complexity of machine somehow becomes conscious in some way. I don't know how you test that even. So there's all kinds of questions there and I'm not an expert in that. I really couldn't
- 39:01 Debate a philosopher who has expertise in that to any significant extent. But in terms of achieving AGI, I don't know why you would have to have consciousness in order to achieve AGI. If you think about the tremendous things that we've been able to do with next token prediction, especially now that we have systems where that next token prediction is kind of happening behind the scenes and it's being double checked, it allows us to have very robust, highly accurate responses. We can trust our agents to do more and more and more things, more complex tasks, no more so than in code. I mean, it's absolutely mind blowing what you can be doing with tools like CloudCode and OpenAI's Codex and the Gemini CLI. So yeah, I don't think that consciousness is necessary. I think that we can take inspiration. Any of the AI systems, large language models that we have today involve deep learning, which involves artificial neurons.
- 40:05 So these are an algorithmic representation, a very, very simple algorithmic representation of how a biological



brain cell works. And when we scale that up very large, we get these amazing capabilities, like I was just saying, like cloud code and all the cutting edge things that you can do with AI today fundamentally involve this artificial neuron at its heart, but that artificial neuron algorithm is such a simple, simple, simple representation of the way that an actual biological neuron works. And so AI researchers have been for decades since at least the 1950s have been taking inspiration from the way that biology works and we can make more complex artificial neuron algorithms. We can come up with systems that are inspired by the way that the human brain works or the animal brain works in terms of its structures. So if you think about things like the hippocampus that's there for memory, where you think about the cortex that is specific to the higher level thought that we have as humans, if you think about the cerebellum, which is specialized for motor tasks, you could take inspiration from any of those structures and how they're connected in a human brain or in an animal brain, but that isn't necessarily going to be the right way to build a better and better machine intelligence because fundamentally you're working with a very different kind of thing with machines we have no limit to how much we can scale up compute in a way with machines that we can't with a human brain.

41:46 You can't like hook a bunch of human... Well, we don't have a way today of like hooking a bunch of human brains together to be getting even more power and there's things move at photons move at the speed of, well, electrons today, but photons in the future in computing move at the speed of light and that isn't something that happens in our brain. Things move much more slowly. They're moving at a chemical space. Things are moving more slowly. It takes tens or hundreds of milliseconds even for your brain to do a simple perceptual task, forget some kind of cognitive task, whereas machines can move much faster. And so there's different constraints in a



# SuperDataScience

biological system versus a silicon system. And so I think while we can take inspiration from the way that the biological systems work, there's also opportunity and limitations in silicon that we don't have in biology.

42:38

All right, that's it for today's. In case you missed an 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.