

**SDS PODCAST
EPISODE 775:
WHAT WILL HUMANS
DO WHEN
MACHINES ARE
VASTLY MORE
INTELLIGENT? WITH
ALEKSA GORDIĆ**



- Jon Krohn: 00:00:00 This is episode number 775 with Aleksa Gordić, founder and CEO of Runa AI. Today's episode is brought to you by Ready Tensor, where innovation meets reproducibility.
- 00:00:09 Welcome to the Super Data Science Podcast, the most listened-to podcast in the data science industry. Each week we bring you inspiring people and ideas to help you build a successful career in data science. I'm your host, Jon Krohn. Thanks for joining me today. And now let's make the complex simple.
- 00:00:44 Welcome back to the Super Data Science Podcast. We have an unbelievably intelligent and fascinating guest for you today, that's the AI juggernaut Aleksa Gordić. Aleksa is founder and CEO of Runa AI, a startup focused on building multilingual LLMs. On the side, he's an online educator that has built a community of 160,000 people in the AI space, including through his AI epiphany YouTube channel. Previously, he was an AI research engineer at Google DeepMind in London, and a machine learning software engineer at Microsoft. He holds a degree in Electronics and Computer Science from the University of Belgrade in Serbia. Today's episode contains tidbits here and there that will appeal primarily to hands-on machine learning practitioners, but it mostly should be of great interest to anyone. In this episode, Aleksa details why multilingual LLMs provide so much value despite the cutting-edge LLMs like Claude 3, Gemini Ultra, and GPT-4 supporting so many languages. He provides his frameworks for entrepreneurial success and for effective self-directed learning. His analogy for how humans are born is a checkpoint of a Bayesian model that's fine-tuned with a reinforcement learning from human feedback. And he opines on what it will take to realize artificial super intelligence and what that could mean for human society when it arrives. All right, you ready for this exceptional episode? Let's go.



- 00:02:11 Aleksa, welcome to the Super Data Science podcast. I can't believe you're here already. Before I even hit the record button, we've had such amazing conversation. I was like, oh man, we've got to just get this record on, everything we just said should already be on.
- Aleksa Gordić: 00:02:23 We should have filmed it, man. Best conversations always have off-cam, so just film the whole thing and then we can edit.
- Jon Krohn: 00:02:31 Well, we'll do our best here. We've got tons of amazing questions and topics that our researcher Serge dug up for you. Where in the world are you calling in from today, Aleksa?
- Aleksa Gordić: 00:02:39 I'm currently in Serbia, Belgrade.
- Jon Krohn: 00:02:41 Nice.
- Aleksa Gordić: 00:02:42 That's Europe for Americans.
- Jon Krohn: 00:02:46 We will be talking about Serbia more actually and the general Balkan region. So yeah, get your maps out Americans. And so we were introduced by Ken Jee, you were on the Ken's Nearest Neighbors podcast, which is a great show. Ken, an incredible content creator and leader in the content creator community really. I was delighted that he introduced us and it's unreal to have you here. Let's get right into the content. So you left Google's DeepMind team, which is arguably and in my view, still the most prestigious AI lab in the world, even though you've left. They still-
- Aleksa Gordić: 00:03:31 Maybe a bit less so, a bit less so.
- Jon Krohn: 00:03:32 A bit less so, but there's still-
- Aleksa Gordić: 00:03:33 Used to be really good.

Show Notes: <http://www.superdatascience.com/775>

- Jon Krohn: 00:03:35 Somehow scraping by without you. And so you've been working on your startup Runa AI, R-U-N-A, which is a full stack generative AI platform for language specific efficient LLMs for government and for companies. But specifically the niche that you've carved out there is multilingual or non-English large language models. So tell us a bit about that, how it's going and why we need these non-English specialized LLMs when it seems like to me, as a primarily English speaker, when I'm using the big LLMs, when I'm using Claude 3, when I'm using Gemini, when I'm using GPT-4, it seems like they're competent in other languages already.
- Aleksa Gordić: 00:04:18 That was a mouthful, thanks for clearly explaining what my startup does. I would just say that it's not multilingual per se in the sense that they actually create dedicated LLMs that target usually one language, but it's mostly bilingual. So it's like English and one target language that's underserved.
- 00:04:35 And the reason I'm doing this is if you take a look at some of those LLMs you mentioned, like take whatever, Claude 3 or GPT-4, they do support... They're multilingual in nature, but there's a tail distribution in the sense that English is very nice to represent and very performant. Then maybe Spanish and French and German are really good, but they're already like two [inaudible 00:04:58]. And then it starts reducing the quality of the outputs rapidly as you go to lower resource languages. And that's the reason why I'm tackling this space, because I see an unfulfilled need. I see something that's hopefully going to be impactful. And I mean it already was with the stuff I've done here in the Balkan region for Serbian, Bosnian, Croatian languages. So yeah, that's briefly what we're doing. Building LLMs for underserved languages and trying to build cool applications around them.

- Jon Krohn: 00:05:31 Awesome. And you also have been open sourcing an initiative called the No Language Left Behind initiative, open NLLB. And so that sounds like it's related as well. So it sounds like you were speaking about already with your commercial initiatives tackling slightly underserved languages. So you've mentioned major languages are well-served by LLMs, English, French, German, that kind of thing. But this initiative sounds like it's going even further and it's saying not slightly less known languages like Serbian and Croatian maybe, but it sounds like No Language Left Behind is talking about relatively rare languages that maybe only a small number of people speak.
- Aleksa Gordić: 00:06:15 No, not necessarily. So let me briefly give a context for those of people who didn't hear about the project, which I guess is most of your audience. The project was where the whole current idea was started. I basically wanted to open source Meta's No Language Left Behind Machine Translation System, which supported 202 languages, across every single direction. So that's 40,000 plus directions. Because it was only for non-commercial use case. So I want to build something that's open source that supports the people can use commercially. And basically through that work I pivoted towards building an LLM instead of machine translation system, for Serbian initially and then expanding from there.
- 00:06:58 But the thing, because it's a startup and it's a for-profit startup, I can't really be focusing only on very, very small languages because I won't be able to sell anything. So you still have to make a trade-off there. And so let's say I'm focusing on everything that's not supported by Mixtral and GPT-4, and... Maybe that's a better explanation than very, very underserved.
- Jon Krohn: 00:07:24 Yeah. So Aleksa, so your first model with Runa AI, as far as I know, the one that was first publicized was called

YugoGPT. So it was built specifically to understand as we said, there would be more Balkan countries here. Serbian, Croatian, Bosnian, and Montenegrin were the official languages of four neighboring countries that were once part of the same country, Yugoslavia. They were at war in the 1990s. And so not to resurface old wounds, but do you think things like LLMs, things like YugoGPT can bring communities closer together and prevent conflict?

- Aleksa Gordić: 00:08:00 I like to think they do, but then again, people who make those decisions probably don't care. So it might be just my bubble. But definitely what I've observed already is that scientists and just people and communities on LinkedIn and those professional networks, that's where that connection definitely happens and cross-collaboration happens. But then again, scientists do collaborate by design, by default. So as I said, it's not really a big progress. Unless you see politicians changing something because of my LLM, I can't say I made such an impact. But there will be a bit of a stretch for me to say I had such an impact. We'll see.
- Jon Krohn: 00:08:36 Yeah, I guess so. That could be true. Well, you yourself, you speak five languages and you're strongly interested in human languages alongside your AI work. I know you've spent time deliberately studying new languages. That's something that you push yourself on. Is that related? Is that interest in languages in general in your normal day-to-day life? Is that influential in what you're doing now?
- Aleksa Gordić: 00:08:58 Definitely. I mean that's something that started in the childhood, or more precisely it started in high school actually. So beginning of my high school, I got really bad grades in German, and then I started just self-educating myself. And there was also it aligned with my other self-improvement efforts, including calisthenics and whatnot. And then I became really good at German because I

started ferociously reading books in German language. And I read, I think in a span of a year and a half, I think it was 11 books. I have it on my blog somewhere. And the last one was like Hermann Hesse, which was the German, not Nobel Prize, but a famous novelist. I don't know what's the term for the award he got. So that's kind how it started. And then it's expanded in Spanish, Portuguese because I spent some time in Brazil in a fraternity with 11 Brazilians and 50% of them didn't speak English. So I literally had to speak Portuguese and I already spoke Spanish, so that helped. So it kind of skyrocketed the learning.

00:09:59 But yeah, there is a passion for languages I had. And the thing is, I had it more when I was younger. Because in the meanwhile I just realized, okay, first of all, it's very easy. So what I mean is the marginal difficulty, if that's the correct term, is not as high as when you're learning your second language. So when you have your fourth, learning fifth from fourth is not that big of a deal, especially sixth from fifth. So it's not that challenging anymore. And also if you're not using it, you are kind of losing it. Not totally. If you ever learn the language, you will not forget it completely, but it does go down the fluency.

00:10:38 So because of those two, and given my current priorities of building a tech company and everything I'm doing in AI, I don't really have the time to just constantly polish those skills. And also I'm bullish on machine translation systems. I think in the future you will know your own language and you will have some type of universal language. And I think that will be probably ideal outcome because language is very important for preserving the culture and the history. So I definitely am very bullish on, we cannot allow that everybody just speaks English and we forget our languages. Not because I'm a nationalist or anything like that, it's just because I like evolution. I like

all the artifacts that culture produced through language need to be preserved in some way, and then we have to be bilingual so that we can actually be efficient and effective. So I'm also pragmatist. But there are kind of those two tensions that you have to balance out, if that makes sense.

Jon Krohn: 00:11:26 Yeah, with somewhere between 75% to 90% of STEM papers being published in English and most technical books being published in English, although we have the political difficulties that prevent LLMs maybe from preventing war and conflict, that does seem like the area that you just described where you have this lingua franca of English across STEM globally, but people don't need to grow up speaking English to participate in that anymore. You can get effectively real-time translations of papers. You could stay fluent speaking most of the time in your home language, whatever that is, but be having instant translations, probably even things like you could be listening to this podcast or watching a YouTube video in the near future where everything is being translated in real time. And so even though English is still used behind the covers, I guess, you don't necessarily need to be exposed to that as the reader or the listener.

Aleksa Gordić: 00:12:30 I agree that's a nice future to have, honestly. I also would argue that there is a value to be had of learning a second language, at least one. Because that's where this effort and challenge interconnect, so in a sense of some type of a cognitive exercise. But then again, if I made that argument and then extrapolated, people were very against technology precisely because of that. And so I don't want to impede progress. Because okay, we now don't have to remember phone numbers or we don't have to do some calculations because we have a calculator. That doesn't mean we got stupider, that just means we are focusing on different stuff. So I think this is going to free us up for doing something else. And that's how it always was. So I

don't think it's going to be a big deal even if we don't have to learn any language other than our own.

Jon Krohn: 00:13:21 Nice. Yeah. It's a great vision and I agree with it for sure. It also it, speaking more languages, expands our cognitive abilities beyond language. Just as learning musical instruments or probably things like yoga, any of these kinds of things that broaden your horizons. But I think languages in particular have a big impact on our capacity to learn and remember concepts in general.

Aleksa Gordić: 00:13:49 One thing I would say here is that there is additional thing that people maybe don't appreciate if they are not multilingual, if they're not polyglots, and that's that every new language teaches you a bit different perspective. So not just that you're forced to actually get to know that culture a bit more. It's that some languages have different concepts of how you represent time, how you represent space, and that helps you later when you're even problem solving. It helps you problem solve. So it's not just rote learning of new words, it helps you learn that some people when they think of future, they don't see the future ahead of them, they see the future behind them.

00:14:26 And that's also intellectually equally good representation because what's behind you you cannot see, what's ahead of you, you can see. And the further it gets, the less you know, which is exactly how past and future work. So it's equally good representation of how time works, even though for us coming from the Western world, for us, future is ahead of us. We are going ahead, past is behind us. But both are equally valid. And so that's one of the things you learn and discover as you're learning new languages, those weird mind shifting exercises, so to speak.

Jon Krohn: 00:15:03 Research projects in machine learning and data science are becoming increasingly complex and reproducibility is

a major concern. Enter Ready Tensor, a groundbreaking platform developed specifically to meet the needs of AI researchers. With Ready Tensor, you gain more than just scalable computing, storage, model and data versioning, and automated experiment tracking. You also get advanced collaboration tools to share your research conveniently and securely with other researchers and the community. See why top AI researchers are joining Ready Tensor, a platform where research innovation meets reproducibility. Discover more at readytensor.ai. That's readytensor.ai.

00:15:44 Next thing you're going to tell me is that there isn't a right side of the road to drive on, it could just be either side.

Aleksa Gordić: 00:15:52 Oh no! Just not that. Yeah, I think we should definitely uniform... Metric system for the win, and just pick a side of the road, whatever it is, just pick a side.

Jon Krohn: 00:16:00 It's crazy. There's actually, this is a complete tangent unrelated to a data science podcast, but there are places in the world where right side of the road and left side of the road meet, where they have to have some kind of solution. And so there's weird cloverleaf things that happen when you get off a highway to switch so that all of a sudden you switch in this big cloverleaf from going on the right side to the left side.

Aleksa Gordić: 00:16:23 Crazy, crazy. I know in London there is a small hotel where that's the only place in the whole of UK, I think, where you drive on the right side or something like that.

Jon Krohn: 00:16:32 That's funny.

Aleksa Gordić: 00:16:33 Anyhow, just trivia.

Jon Krohn: 00:16:36 Speaking of technical challenges like that, I don't know if you want to go into, I don't know if you can go into

without divulging anything proprietary, but with your approach, with YugoGPT or the kinds of things that you're doing at Runa AI, are there things you can tell us on air that are exciting about the way that you're doing it and novel and pushing boundaries?

- Aleksa Gordić: 00:16:55 Until I publish a paper, I'm afraid I can't go into too much detail.
- Jon Krohn: 00:16:59 Makes perfect sense.
- Aleksa Gordić: 00:17:00 Yeah. Yeah.
- Jon Krohn: 00:17:01 Nice. Well, let's talk about your entrepreneurial experience in general then. So having experience with tech giants in the past, like Microsoft, like Meta, like Google DeepMind, and now leading your own startup, Runa AI, what key lessons would you share about the opportunities versus the challenges of entrepreneurship in AI relative to being at one of these big established tech companies?
- Aleksa Gordić: 00:17:24 I'll preface it by saying that until I make it, really make it in entrepreneurship, I'm not the right person to be giving advice. But then again, you are the best person to give advice to somebody who's one step behind. So from that perspective, I think I can give a couple of advices.
- 00:17:39 So first of all, while I was still working at Microsoft and DeepMind, I was constantly doing something on the side. So that makes me entrepreneurial over the past many years, not just over the last year since I left. It was literally I think 20 days ago I left, last year. And now when I say entrepreneurial, I mean one of those was growing a community, a sizable community, YouTube channel, LinkedIn community, almost 100K there, 90K to be precise. Then Twitter, Discord. So that whole community kind of thing could be treated as a business

and it is a business because I do get money from sponsors and stuff when I do those.

00:18:14 So that opened me up to this whole world of entrepreneurship. So Quantum was my first thing there. And then I did attempt a couple of projects in apps. So even before I left, I think in 2016 or '17 while I was still studying, I made a small Android app. And we pushed it to play store, but we had zero understanding of how the markets work, how competitive it is. And so what happened is we expected some type of skyrocketing trending news or whatnot application. And because it was like a meme app, so I thought it might cut and be viral. But literally it was so underwhelming and I was like, okay, you have a lot to learn and yeah, let's maybe first get some credentials and knowledge before you do something else. That was kind of how I was thinking about it. Yeah, I mean, I can give you some general piece of advice, but I'll let you steer the conversation wherever you want it.

Jon Krohn: 00:19:08 Yeah, it seems like in general, something that you just pointed out there, there is a concept I think with a lot of early entrepreneurs that if you build it, they will come. If you build this great app, people will just love it and it's going to take off. And that isn't reality. It's very hard to get people into your app and to make it sticky, which is a-

Aleksa Gordić: 00:19:31 Yeah.

Jon Krohn: 00:19:31 Yeah.

Aleksa Gordić: 00:19:32 That's the thing with probabilities, probability wise, you are very unlikely to have such a situation. But I would argue that there were companies, like take Facebook, it did happen to them, it did come to them. But it's just survivorship bias. You see a couple of those that make it. And for them, it was actually true that it came to them.

Because they were just in the right space, right time, right place, right niche, right product. It's not that Mark deliberately has done a market survey when he was 19 years old and was like, "I'm going to do that and that because it makes sense." No. He was like, "Oh, this is cool and fun, and I like to see other people's profiles. People are curious about people." It turned out it's just a very popular product. And so it does happen, but you cannot plan for it. If it happens, kudos to you, you better get that opportunity and make the best out of it. But if it doesn't happen, then you just have to grind like what most people do have to do.

- Jon Krohn: 00:20:26 Another thing that came up in my mind, a great question related to what you mentioned. So you talked about how this app, this meme app, it was your first foray, didn't work out. You decided that you needed to learn more before tackling another one. How do you know when it's the right time to drop a project and start something new or develop some other skills? How do you know as an entrepreneur when you're just continuing to throw now good money and good time after bad?
- Aleksa Gordić: 00:20:57 Yeah. Well, on one side you'll never be ready. So from that standpoint, if you're looking to minimize the risk, you better just stay at your big tech job. On the other hand side, you can feel it. So there is something, we live in a hyper-rational society, especially us in the tech bubble. But there is a gut feeling, there is intuition, there is a tension, there is the fact that time is ticking and you only have so many rolls of dice before you potentially make it or fail. So the sooner you start the better. But then there is a tension that you have to start as soon as possible, but then there could be too soon potentially because you end up in a local optimum spending some of the best years of your early twenties and potentially not learning that much.

00:21:48 I mean, you will always learn something. You can always justify in hindsight, "Oh, I learned that and that." but it's not a matter of whether you learned, it's the opportunity cost. If you picked a different path for yourself back then when you made that big decision, would you be much better off right now than by doing what you're doing? And that's the thing that you have to keep that in mind. Opportunity cost is a very important concept.

Jon Krohn: 00:22:13 Yep. Well said. It's interesting that it seems like you have this well honed sense. You talked about this intuition of when it's the right time to move on. Before we started recording, you talked about 2X multiples. Is that something, is that kind like a general rule of thumb that you would at least apply in your endeavors that if you're not getting 2X growth on something, maybe it's not even worth pursuing, you should drop it in and-

Aleksa Gordić: 00:22:35 No, I think that's the optimistic of a goal. I think that sometimes when you really believe in idea, you should grind. I mean, take a look at the Nvidia stock chart man. I rest my case. It's just like it was not growing 2X, it was literally, if you zoom out right now, 25 years, it was literally flat. From the macro perspective, flat. And then all of a sudden 2017 comes and you see exponential growth. So definitely do not follow that advice. It depends on what you're doing. So for content creation, I definitely think you do need to have some type of exponential growth. Obviously exponential is always turned into sigmoid functions because there is only so many humans on the planet, so it has to saturate. But in the early days, you definitely do have to see some bigger traction. But also sometimes you have to suck through the local optimum where you're not growing as much until a breakthrough happens and then you can explode.

00:23:38 So there are all of these tricky local optimals that you have to endure and believe in yourself. So there is no

really single piece of advice other than, yeah, you have to think through your particular situation and problem solve your way and balance between being stubborn and persevere versus being stupid and just pursuing a dead end. And that's the tricky thing about entrepreneurship because everybody is in a different point of the search space, so to speak. And so none can give you the precise advice you need other than you or maybe somebody who really knows you really well and can help you get unstuck from there, if that makes sense.

Jon Krohn: 00:24:20 So talking about projects, and related to social media that you might've moved on from, you launched a startup called Ortis to ask questions from YouTube videos. And that appears to be something that you're not pursuing anymore, is that right?

Aleksa Gordić: 00:24:33 Yeah, that was the initial idea I kind of pursued when I left DeepMind. I wanted to build AI first video platform. So it was supposed to be a new AI first YouTube, so to speak. That would be layman's explanation of what it was supposed to be. And then I started it as an MVP of just like, let's start, instead of starting to build the info and everything, before I even know people want this, I just built an MVP. And there was a Chrome extension for YouTube where people could type in a question, you get an answer about that particular video and you get relevant time stamps. So if you ask, hey, when did Sam Altman mention 7 trillion parameter GPT-7 model? It's going to find you exact moment in the video where that is and give you the timestamps and give you explanations. And even though it sounded like a very cool idea in theory, it turned out people don't really care about that. Most people won't interact, when they're watching videos they're passive. Most people are passive. So if you rely on them being active, you are immediately serving some other type of audience. That was one of the realizations I could have made even before I pushed this out.

00:25:42 But it turned out to be one of those projects where everybody was looking applauding you on social. It's so super cool. And then you look at metrics and those folks are not using it, and that happens a lot. I noticed how empty a lot of the social media reactions are. Ultimately the only thing that matters is does somebody care about your product? Are you making somebody's lives easier or better? Or as YC puts it... What's their saying? Just make something people want. It's as easy as that. One realization for me is you cannot work around that having by being like AI influencer or whatnot. I hate that word, but for lack of a better clear word, concise word, I'll use that one. But yeah, it doesn't matter if you have 500K followers on whatever on platforms. If you build a [bleep] product, people are... Initially you're going to spike it up, the same that happened with Threads with Zuckerberg, but then it's questionable whether they'll keep on using it, and that's a function of how good the product is.

00:26:44 Now you lose all of that initial distribution privilege. You do have that distribution type of advantage compared to somebody who doesn't have an audience, but that's where it stops. It doesn't mean you're going to succeed.

Jon Krohn: 00:26:56 Right. And so as cringey as the term AI influencer is, if we had to use that term, I could describe it for you for sure. You have 50,000 subscribers on your YouTube channel, The AI Epiphany. How do you balance that with your entrepreneurial ambitions? Does it complement it? I mean, you just talked about one advantage there that at least you get an audience to check out something new that you build. It doesn't guarantee that they're going to use it doesn't guarantee that they're going to stay. But at least you have a waiting audience to be trying something out that you create. So that must be one advantage. What other reasons do you have for content creation or why do you do it?

- Aleksa Gordić: 00:27:31 So YouTube for me over the last year has mostly been me uploading some talks. I've been having, like live talks on Discord with various AI researchers, engineers, CEOs, I had Jeremy Howard, I had folks like that, like Thomas Wolf the CSO of Hugging Face. So from that standpoint, it's very, very low effort. I just have to film something during the talk and I use that learn and meet people and interact, and then I just upload it to YouTube. From that standpoint, it's not really a big of a time drain. That's why I can really easily complement it with what I'm doing in my startup. There will be a TLDR.
- 00:28:14 But before I was actually filming longer videos, I was famous for making these hour and a half, two hour long videos where I go through a paper, through all the formulas, all the explanations, or step through the code base literally with a debugger, NVS code, line by line and explain everything. So those types of videos take much more time and it's not justified for me to do it anymore because I could be using time better in a different way.
- Jon Krohn: 00:28:40 All right. We're going to move on to another topic soon, but last one here on entrepreneurship. In a recent interview you mentioned turning away offers from big tech companies to do your own thing. You've already talked a little bit in this episode about how if people don't want to have risk, big tech is a safer place to go. For you personally, why is it obvious that you should be taking these bigger swings, taking these bigger risks, trying out different products? Why is that the path for you as opposed to a potentially more secure path?
- Aleksa Gordić: 00:29:15 Because first of all I'm 29. When am I going to do it if not now? In my 20s or early 30s or even 40s, there is no moment when it's too late. There's so many success stories of people with even 50s or 60s and building cool companies, although not tech companies per se, it's a bit more competitive there.

00:29:32 I mean, how I see it is since I left I've been building all of this and open sourcing some of these models. I got a lot of cool offers that I would probably not have gotten if I just stayed at DeepMind and I learned much more because I have complete agency, I can do whatever I want, I can build what I want, I can hire whom I want, I can focus on whatever I want. So from that standpoint, there is a lot of agency and learning that I had with hiring people, dealing with teams, talking with VCs. All of these things help me better understand how to build a product and have a better holistic understanding and view of the business world. As well as learning a lot of cool new technical stuff, so it's not like I'm losing my edge and becoming less employable. On the contrary, I would say I'm much more valuable now to join some company because I actually know what's fluff and what actually matters and what you have to focus on. Everything I said about distribution and some realization there on attention of the product versus the quality of the product and what you want to focus on and the importance of timing. And so all of those meta things, it's very hard to describe verbally sometimes that I learned over the past year.

Jon Krohn: 00:30:57 Eager to learn about large language models and generative AI but don't know where to start? Check out my comprehensive two-hour training, which is available in its entirety on YouTube. That means not only is it totally free, but it's ad-free as well. It's a pure educational resource. In the training we introduce deep learning transformer architectures and how these enable the extraordinary capabilities of state-of-the-art LLMs. And it isn't just theory, my hands-on code demos, which feature the Hugging Face and PyTorch Lightning Python libraries guide you through the entire lifecycle of LLM development from training to real-world deployment. Check out my generative AI with large language models hands-on training today on YouTube. We've got a link for you in the show notes.

Show Notes: <http://www.superdatascience.com/775>

00:31:39 Makes a lot of sense. For me, in addition to entrepreneurial stuff, the content creation itself, having these conversations with you, doing the research on you, understanding what you're doing, this is also, it's a way for me to have a lot of fun while I'm learning. So I totally understand everything you're saying there and I couldn't agree more that by going off on your own trying to do things, those skills that you learn make you more attractive if you want to go back to big tech or whatever job afterward anyway. So it's one of those situations where who dares, wins. If you dare, you can't lose. Even if on the surface, okay, the big swings didn't work, you didn't have this product that really took off and you became Mark Zuckerberg, that kind of survivorship bias example that you gave earlier. But nevertheless, you learn so much more when you're in that kind of situation.

00:32:39 One thing you mentioned there is you said you can build what you want, you can hire who you want. When you're on your own like you are now, do you have difficulties handling the infinite number of options that are available to you and that agency that you have that is one of the best parts of being on your own can also be one of the most intimidating? It doesn't seem like it's an issue for you. It seems like that you thrive on that infiniteness of possibility?

Aleksa Gordić: 00:33:15 That's a good question. Honestly, I would have to think about more hardly about where I stand with regard to that question. It is definitely sometimes overwhelming when you have all of that optionality, I agree. But then again, after having spent so many years at Big Tech I was thirsty for this type of agency where I can just do whatever I want and get that type of a win, so to speak. And so worst case, I joined some AI lab and then in a couple more years I try again. But this time I actually failed once, so that's actually an asset and something that

puts me in a much better position because I know now what to look for.

00:34:01 For example, one of the biggest mistakes I made during my big tech career is that given that I knew that my goal is to start a startup eventually I didn't focus more on people and building relationships and looking for that co-founder while I was working there. Explicitly always keeping in the back of my head, hey, this is a person I might recruit or have as a co-founder next day. I never thought about that explicitly and I was not deliberate about it. And I think that was by far my biggest mistake I made so far in my career.

Jon Krohn: 00:34:33 So that's actually something right now you have on your LinkedIn profile that you are looking for a technical co-founder. So at least at the time of us recording this, your current role is listed as solo founder of Runa AI. And then in parentheses after your job title it says solo founder, open bracket, actively looking for a technical co-founder, close bracket. So that seems to tie in here. And so, one quick thing is it seems like you are already a technical co-founder, so it's interesting that it sounds like you are looking for your CTO to compliment you as CEO potentially. And so you can fill us in on that. And then yeah, let us know what you're looking for. Maybe your technical co-founder is listening to this podcast right now.

Aleksa Gordić: 00:35:20 That'll be amazing. I mean, first of all, why waste a good real estate? I know how many people are visiting my LinkedIn profile every day and actually my previous co-founder, we split away since then, but he found me through my LinkedIn. And so I'm very well aware of the views I'm getting there and so I'm just using it. So that's why I have that message there.



- 00:35:44 And then the thing is I talk with a lot of people and a lot of them have amazing backgrounds. I even talked with some of my ex-DeepMind colleagues. The thing is you cannot compensate that easily for the lack of not having spent the time together in the same room and building for multiple months. The best kind of alternative to that is to fly over wherever that person is or vice versa, and you spend some time building together before you commit. And that's what I've done with my previous co-founder. But despite that, it's still turned out we are not the right match. And since then I realized the reason... You asked me about the technical part. So obviously I'm very technical. I'm a technical person. I worked as a software engineer initially at Microsoft, then as a machine learning engineer and finally as a research engineer at DeepMind. So I had to pass all of those hard tech interviews and I've demonstrated by building all of these products in the open that I know how to build really anything. So it's not about that.
- 00:36:44 It's just that when you're building a startup, I'm talking with customers, I'm doing this, I'm talking with VCs, I'm networking, I'm doing a ton of stuff. I be just a cog in a system building a single thing. I'm very breadth research so to speak. And so I need somebody to compliment me and be the chief science officer or chief technology officer or whatnot. Yeah, that's the brief explanation there.
- 00:37:12 But also if you take a look at YC Advice, everything they say there is much better to start off with technical co-founders and learn sales and business stuff because it's easier to go that direction than to go from business to learning engineering and learning that type of analytical thinking. Now, I wouldn't 100% say that's always the case, but it's one of those things it's ballpark correct. I think that's ballpark correct thing to say. Unless you're truly, truly introverted, you don't know how to handle anything, then you'll probably never be able to learn

about business or sales or those things. But if you're just like a regular human being and you have engineering background, you're going to be easier learning desktop than vice versa.

- Jon Krohn: 00:37:53 I agree with you on that and that is a great concrete tip for our listeners there. Lots of great concrete tips for people to read on the Y Combinator Blog.
- 00:38:02 So you mentioned in there we're now moving on to another topic, but you provided a perfect segue, which is your open source contributions. So you talked about how you have made tons of projects that you've shared on your GitHub. How do you decide which projects to work on and open source? I guess this kind of ties into earlier I had the question about the infiniteness of possibility, and so this is I guess almost a follow-up question on that. Where with open source projects, there's again an infinite number of things you could do. How do you pick from all the things out there? Is it just kind of, do you have that intuition that, wow, this thing is the most exciting thing right now, I can't imagine working on anything else, I'm just going to do that? Or, is it more methodical?
- Aleksa Gordić: 00:38:49 It's kind of both. I did do some survey last summer of the things across very small modalities and seeing where the gaps, so there was some systematic approach there as well. With YugoGPT in particular I knew that basically it's going to explode here in the Balkan region just because there is this big thirst and knowledge about ChatGPT and lack of local languages and LLMs being all the hype yet zero good models for these languages in particular. And so I had the prediction going on back in October or November and then my prediction turned out to be completely correct somewhere I think mid-December or late December last year when I opened sourced YugoGPT and we had a crazy media attention across hundreds of portals, newsletters, and television shows inviting me

even though I said no to 100% of the new shows because I literally didn't have time to do that stuff. I found it a big distraction. But I did write out a couple of a blog. I got some written interview questions and I did reply to those and I was even covered by the national television website or what not. Anyhow, it turned out to be a big, big hit here in the local region.

- Jon Krohn: 00:40:20 Nice. Yeah, so it sounds like you do kind of have an intuition around these things. Maybe through all of your reading, that probably helps, maybe all the language speaking and the other activities you have to clear your mind, it comes to you this vision of YugoGPT, this is going to be huge. Let's go.
- Aleksa Gordić: 00:40:43 In this particular example I would say it was less about intuition, more about actual explicit knowledge. So when I say intuition, it's something you can't really articulate, you don't have any concrete evidence, it's just a bit more abstract and you have to make these decisions in these highly dimensional search spaces and you just have an intuition for where to go. Maybe similar to what Monte Carlo search tree does in Alpha Zero or something like that. You have some type of a value function and you estimate the value of this particular path and you go that direction. But that happens completely outside of your explicit conscious thinking where you're having these monologues with yourself explicitly. No, it just happens. You kind of feel it and then you go that direction. That's the best way I would describe it, knowing how these models work. It's a value function, internal value function.
- Jon Krohn: 00:41:36 All right, so we've already talked in this episode about your entrepreneurial background, Runa AI, open-source contributions, all of these things relate to AI. Let's dig into AI specifically. You have a very strong background in here, as you talked about. There are issues around things

like the energy demands of large language models that these could pose sustainability challenges. Do you foresee solutions that will allow us to have LLMs, transformers be more efficient and environmentally friendly going forward?

Aleksa Gordić: 00:42:07 Oh man, I'm a big believer in technological progress and historically just extrapolating from the history, there is no reason for us to be that worried about those aspects because I mean, in five years we'll be running some of these huge models on very small devices and they will not be wasting that much energy. So from that standpoint, I'm not that worried about the environmental problem. On the opposite, I would say the only way to solve the environmental problem is for us to just keep on expanding our technological capacity as a civilization and get with better creative solutions to some of the pollution sources we currently have. One would be imagine us building fusion reactors, that would completely change the way we... We would go on a different... I forgot the name. I think it's Kardashev scale, right? We would go to the next level of the Kardashev scale. We would be able to basically use all the energy coming from sun or equivalent energy by building our fusion reactors.

00:43:16 So from that standpoint, I just think we need to accelerate. And by saying that I am not saying I subscribe to the EEC like dogma. I just think I am very technological progress and that's the best way we can solve not just energy issues, but also cancer and all the other issues that are plaguing humankind, honestly.

Jon Krohn: 00:43:39 Ready to master some of the most powerful machine learning tools used in business and in industry? Kirill and Hadelin, who have taught millions of students worldwide, bring you their newest course Machine Learning Level 2. Packed with over six hours of content and hands-on exercises, this course will transform you

into an expert in the ultra-popular gradient boosting models, XGBoost, LightGBM and CatBoost. Tackle real-world challenges and gain expertise in ensemble methods, decision trees, and advanced techniques for solving complex regression and classification problems. Available exclusively at superdatascience.com. This course is your key to advancing your machine learning career. Enroll now at superdatascience.com/level2. That's superdatascience.com/level2.

00:44:21 Yep. I am in the same boat as you, and that's part of why I host this show is relentless techno-optimism. We've got to stop spending money on bombs and get that into nuclear fusion and into agriculture and all kinds of other solutions that could be making the world a better place, education. And speaking of education, to give you maybe a challenging question here, it'll be interesting to hear your answer. So Yann LeCun recently argued that in four years a child has seen 50 times more data than the largest LLMs today trained on all the texts publicly available on the internet because the data bandwidth of visual perception is about 1.6 million times greater than the data bandwidth of written or spoken language. So he goes on to say most of human knowledge and almost all of animal knowledge comes from our sensory experience of the physical world. Language is the icing on the cake. We need the cake to support the icing. End of quote.

00:45:27 So what do you think about this? Do you think Yann LeCun is right? Do you think for achieving AGI pursuing this language approach, language first approach is the wrong route to go down and we should be focused on a bigger sensory experience?

Aleksa Gordić: 00:45:42 So first of all, I don't disagree with Yann, and secondly, I would say that nobody's focusing just on language. If you take a look at what everybody's doing, if you take a look at the recent Gemini model, it's all multimodal, GPT-4

Vision. Everybody realizes if you can get more sources, more modalities of data, why not use it? [inaudible 00:46:01] like sound and doing transcription using Whisper or whatnot. Every single data modality you can get your hands on, you should be using it.

- 00:46:11 Back to Yann's hypothesis. That's something I kind of believe deeply since I got into ML. Every time I hear, hey, like LLMs use trillions of tokens, yet baby can just read five pages of a book and then she knows everything. Dude, you are not counting in evolution. I mean it's so obvious. You are already an artifact in the search space of evolution that took billions of years and that was computation on a very abstract... That's computation. And so you start from a checkpoint. So baby is a base model, baby is a pre-trained base model, and then on top of them you do fine-tuning of your life and then you are also RLHF by the society into doing what's acceptable, what's not acceptable, and that's it.
- 00:47:03 Of course I am falling prey here to the common historical thing, and that's comparing human cognition with what's currently available. People used to compare it with initial analog computers. Before that they were comparing it with mechanical devices. Now we were comparing it with LLMs because that's the closest we've got to AI. Even though we still don't really know how to build general or super general intelligence, meaning better than humans. But we are getting there and I don't see any reason why we shouldn't be scaling up. I agree with Yann in the sense... So the practical consequence of Yann's thinking is that we need much more compute and much more tokens, and we don't need to slow down, on the opposite we need to accelerate there. So that would be my conclusion from that hypothesis, which I think I believe is correct.

- Jon Krohn: 00:47:53 And tokens not just from language, but from these other sensory modalities as well.
- Aleksa Gordić: 00:47:57 Yeah, when I say token, it's the most abstract thing ever, right? Because if you read the vision transformer paper, for them image is basically chunked into patches, and patches are just tokens. So in that sense, a piece of the image is a token, a piece of the speech is a token. Everything can be a token.
- Jon Krohn: 00:48:15 Makes perfect sense. In a Forbes article, you spoke about the second AI big bang being the introduction of transformers that underpin large language models. What do you see... Or do you have any ideas, care to speculate on what the next significant leap in AI technology where that will come from mean? So you talked just now about scaling up, that's one approach. Do you think scaling up on its own, scaling up existing architectures like transformers with more data, more tokens, more modalities, that in and of itself will be enough to say, attain an intelligence that is greater than all humans? Yeah, that was a long question.
- Aleksa Gordić: 00:48:59 Yeah. No, that's a great question. There is something special about efficiency and scale. Again, referring to the common, The Bitter Lesson by Rich Sutton, it turns out the algorithms that get selected as evolutionary, more powerful ones are those that scale with data. They tend to stand the test of time. And so, looking at the current landscape and seeing how many tokens we are leaving back on the table, just think of video, think of YouTube, how many videos is there on YouTube and how much data is not being used. I honestly believe it's not unimaginable for me personally to see the current approaches improving the engineering, improving our infrastructure, having better AI accelerators be GPUs or some of these custom chips. So working on that front, just incrementally, I mean, when I say incrementally, it

kind of goes exponentially, but just keep on getting more compute, more tokens, scaling up is going to get us very, very, very far. Now yeah, whether that's enough to get to general intelligence, I don't know, but it feels like it can. And the best reason, the best explanation I heard was actually from Ilya Sutskever and he said something like this: Why is the next token prediction such a powerful function? So it's just predicting the next token. Why is that so powerful? And why is it the case that it can take us to general AI?

00:50:41 And the reason is, I think, the following story he told, and that's, imagine you have a crime story and you have a couple of suspects in that story. Somebody murders someone, there's a detective. At the end of the story, the detective comes and says, "I know who is the killer. The killer is," fill in the blank. And so for you to do just the next token prediction, the easiest way for you to do that is for you to understand the story. And that's the crazy part about next token prediction, that's why all these properties emerge. Because you're just trying to minimize that loss function to make it easier to predict. And then if you actually find the minimum, the minimum will be, if you understand the story, that's the easiest way you can predict the token and that's it.

00:51:33 And so because of that explanation, I had an epiphany moment, no pun intended to my YouTube channel, hearing Ilya say that so concisely, a single sentence that made me think, "Okay, well I think these guys are on to something." And he saw it much before many other folks. Definitely much before DeepMind folks and the crew, they've done amazing job, in science in particular. But when it comes to scaling and everything that's happening right now, the sole credit goes to those guys. And of course, transformer paper did have a glimpse that, "Hey, these curves seem to be going down." But I'm still mind-

boggled that nobody at Google was like, "Dude, if it's going down, why not push it a bit more?"

00:52:15 I mean, I understand why not because you're in a bureaucratic machine and you have to ask for significant funds. And here you have a startup Silicon Valley mentality and we're like, "Dude, I'm going to make this big bet that transformers when scaled up are going to be amazing." And they've went ahead to do that and the rest is history. So I applaud them for that. And that's not a small thing. You have to believe in something that nobody has done before. It's much easier to replicate later and build up another LLM.

Jon Krohn: 00:52:40 Do you think that approaches like those used for DeepMind's alpha geometry or the rumored OpenAI Q* algorithm where there's a blend of not just next-token prediction, which you could kind of think of in Daniel Kahneman's Thinking Fast and Slow kind of concept. That next token prediction, like you and me just here having this conversation, what word comes out of my mouth? I'm not spending a huge amount of time deliberating on it, it's just flowing out. I'm in my thinking fast mode as we're hearing this episode. And so it seems like there's kind of an analogy there to next token prediction that the leading LLMs today do.

00:53:21 On top of that, when we finish this episode, I could theoretically, I wish this was actually what I was doing next, I could open up a calculus textbook and I could start working through problems. And then I'm not in that thinking fast mode, I have to slow down and deliberate and keep going back over the same information and consider each sequence of my thinking step-by-step. And so yeah, so these kinds of approaches like alpha geometry, like OpenAI's rumored Q* approach, involve this kind of thinking slow, this more step-by-step, deliberative, maybe not even linguistic in the case of

alpha geometry. Do you think that that kind of thinking slow, some kind of mechanism there, in addition to a transformer architecture scale that could provide some gains?

- Aleksa Gordić: 00:54:10 I don't think it's mutually exclusive. I don't think it's incompatible. And by that, what I mean is if you saw, starting from chain of thought to tree of thought to all of those methods where the LLM is having an internal monologue and then using that to output something at the end, that's kind of simulating those types of inner monologues and that's the way we're simulating these internal system two thinking that you refer to, I would say.
- 00:54:45 Now, whether there is potential, I see definitely a lot of potential combining these methods with some types of tree search algorithms, something like Monte Carlo tree search that AlphaZero uses, so those types of things. But we are already doing that and I think we're doing that on top of LLMs. So in that way, I would maybe agree that LLMs are like system one thinking and then what you do on top of LLM, be it like a monologue or creating multiple generations and then having some way to score those and pick the best one, all of that can be happening and can be labeled as system two.
- 00:55:31 This, by the way, reminds me of, I don't remember the guy's name, but he had this theory of a thousand brains, I think. And the idea there is that below your neocortex in the subcortical structures of your brain, you have all of these competing ideas. You're generating thousands of ideas, and those can be maybe LLMs, whatever. And then there is some system on top of that that cherry-picks the best ideas, that one wins and is promoted to your consciousness, and that's the thing you're kind of aware of. But sometimes, and people, I'm not recommending drugs or anything, but people on drugs definitely seem to

display some of these things where they just have a ton of thoughts that they cannot control. And my hypothesis, very layman hypothesis I'm going to say, I'm not a neurobiologist, but my suspicion is that the default mode network gets deactivated and your consciousness drops down into the subcortical structures where all these competing ideas are happening and that's an additional proof or piece of evidence for why I think everything I said makes sense, anyhow, hopefully.

- Jon Krohn: 00:56:39 Yeah, I love that example. I loved it. I loved it. And to give listeners a good sense of for me, an experience without hallucinogens that shows me that same kind of all these ideas happening under the surface that are out of conscious perception but are nevertheless driving your conscious perception is, it's been a while since I've had to write a multiple choice test, but back when I was an undergrad, we still did it with paper and pencil. And something that I would do is if I got to a question where I didn't immediately know the answer, I could say, okay, well of these four answers, I know two of them are wrong. So I'll cross those two out. And then I wouldn't stress about it. I would be like, "I don't know what the answer is between those remaining two, but I bet it'll come to me, and just keep working on the test."
- 00:57:34 And then maybe five questions later or at the end of the test when I come back and come back to this question where I'd crossed off two of the answers, immediately I know which is the correct from the remaining two. And so that for me was always a very tangible expression of that subconscious thinking. I spent zero time, the whole rest of the test I was consumed entirely by the other questions, I didn't put a single instant of conscious thought on figuring out from the question that I didn't know. And nevertheless, my brain was there working on it and figured it out.

- Aleksa Gordić: 00:58:05 I mean, thanks God, because if we were sequential machines, things would be a bit harder. We have all these, you basically spawned a parallel process in the background, and to use some silly CS analogy, it's much more complicated than that, and of course you can do that. And we've known that since antiquity, literally. Wasn't there that, who was it, Archimedes or something was in a bathtub when he realized and he had this famous eureka moment. And I think even listening to David Silver, he mentioned somewhere on some of the podcasts I listened with him, when you're really focused hard on some idea, and then he went to a vacation, he was on a beach just laying down, all of a sudden he solved the problem.
- 00:58:48 And yeah, there is this concept of diffuse and, what's the terminology, convergent and divergent thinking, right? From these courses, like Learning How To Learn from Coursera or whatnot. And that's precisely that. Sometimes you're in a convergent mode where you're just trying, you're straining your prefrontal cortex and then you go into the chill mode where all of these synapses actually start connecting or whatnot, if that's what happens, again, I'm a bit of a layman in neurobiology, and similar to what muscles do, right? It's not like the muscles form during the training, that's when you rip them. And then during the break time when you're resting, that's when the connections are being made. And so probably similar because both are cells and tissues, I wouldn't be surprised that similar things and processes are happening in the brain, as well. Might be a bit different, but anyhow.
- Jon Krohn: 00:59:38 Yeah, I think there might be some mechanical differences, but the analogy makes a lot of sense, for sure. So all right, cool. We've talked about integrating the LLM thinking fast system, system one in Daniel Kahneman's language, with some things that could be more like

system two, deliberate, conscious thought. Some other kinds of approaches that you've spoken about before as potentially being helpful for tackling unsolved problems in AI include the idea of integrating Bayesian learning, graph neural networks, and reinforcement learning. And so I don't know if you want to talk about that any more here, if you have any further thoughts on that.

- Aleksa Gordić: 01:00:22 Well, you digged that up probably somewhere from LinkedIn years ago, because I don't remember when was... I know I posted it somewhere at some point, but it was long, long time ago, so kudos to you doing the research.
- Jon Krohn: 01:00:34 This was in Analytics India Magazine, AIM, in December 2020.
- Aleksa Gordić: 01:00:40 Okay. So as I said, four years ago, literally. Well listen, back then, so what I've done there is very simple, I followed a very simple principle. You have all of these separate areas of ML which don't seem to be converging. So you have graph ML, which is a field for itself, because it never went through the scaling laws, you have LLMs, you have Bayesian with all these modeling probability, you have to model everything as a probability. And so it's just a bunch of disparate ideas and there might be some cross connections across the fields. And that was my guess back then when LLMs were still not all the craze. I don't even remember when the blog was published where it was already May 2020, when GPT-3 was first published. So anyhow, that was my best prediction back then. But it was kind of, I followed a silly principle. I was just like, "Okay, maybe some combination of these would be cool." That's it. There was no deeper insight there.
- Jon Krohn: 01:01:40 I think those were the dark GPT-2 days still. Well, it's great to hear your thoughts on AI, Aleksa, beyond just all this cognitive work that you do. As you've mentioned in

the episode, you are into things like powerlifting and calisthenics. Can you draw parallels on the discipline and the competitiveness that you hone in those kinds of sports, maybe other hobbies that you have, that are helpful to you in AI research and developing and maybe in entrepreneurship as well?

Aleksa Gordić: 01:02:11 I love that question, because I know both of us share that passion. I think I saw some of your lifts and I think I saw a deadlift and you were doing cleans or whatnot, clean jerks, and those are much more from Olympic weightlifting. So for me personally, let me see where I can start. I think my journey started, okay, first of all, basketball when I was a kid, because that's what you're supposed to do if you're from Serbia, you're good at that sport. And then martial arts was my first exposure to this individual game, where in basketball, it's all about team. You don't have to be a peak athlete or anything, you just have to be the best team. Superorganism, so to speak.

01:02:53 And in martial arts, what I learned was the art of self-discipline. I was just like, hey, you got to do these sit-ups and push-ups and squats or whatnot and then I slowly extracted that from that very narrow scope of a training where somebody is doing that for me, I have a trainer who's leading my workout, to me doing that on my own. And I started my calisthenics journey and initially I was just doing push-ups, pull-ups, all of that, trying to max the numbers, and then slowly transitioned. At one point, started going to gym and optimizing for some of the lifts, like bench press, squat, and deadlift. But I wouldn't say powerlifting as much because I never tried to max my one max rep. I was going to having a perfect form deep squat, for example, three reps or something like that. So I never was trying to max the numbers. So from that standpoint, I'm not strictly speaking a powerlifter.

- 01:03:53 But going a bit more meta to the stuff that translates from the sports to everything else I do, I mean obviously one is to set a goal and then execute on it and don't stop no matter how you feel unless you're hurt, literally hurt. So obviously, before a workout you know what you're going to do. You go there, it hurts. Yeah, sure it hurts, but it's supposed to hurt. You push through it and then at the end you rest and your muscles start building up. So that discipline and grit definitely comes from, sports played a role there for sure.
- 01:04:26 And then longer-term planning, as well. Because here you're doing this miserably small amount of work and you're seeing miserably small increments of progress on a daily basis, but then in three months you see a progress. And so you start making, you have this visceral way of making analogies with cognitive space. Where okay, I'm learning today, I am not closer to being Einstein, it doesn't look so, but then three months later, you're definitely better at this. And then even though you are not seeing that progress the same way you can see it in a physical body, you know it's happening. And that gives you that maybe belief, self-confidence or whatnot. So that's maybe a different point that I would make other than just grit and perseverance and those attributes you can hone and develop during your life by practicing sports. So those would be some brief takeaways on how I found value.
- Jon Krohn: 01:05:20 I love it. Yeah, grit for sure. Not stopping unless you're hurt and acknowledging this small progress on a daily scale where it doesn't feel like you're getting anywhere, but taking a step back, you are making a ton of progress.
- Aleksa Gordić: 01:05:36 I would just say one thing. There is actually a big caveat here. There was a con side. There was a con side for me starting with calisthenics and going into learning. So sports teaches you all about execution. So that translates

much better to stuff where you have a charted goal that's completely clear and you just need to execute to get to that goal. But the thing is, that same way of thinking was blocking me from being more creative and understanding and thinking more of independent thoughts. And that's something I've been thinking about much more over the last months: how many people are just recycling other people's ideas and how rare it is to hear somebody say novel things that honestly get me surprised where I can see, "Hey, this person has given a thought to this particular topic." Those are some novel thoughts coming from that person, not just recycling Lex Fridman or Huberman or whatever podcast that person watched.

01:06:36 And so for that type of a thing, you do have to have a bit different mentality. Sometimes it's okay not to do anything for days and be lazy. So that's something that was not natural to me coming from sports. So I have to force myself to just chill, just do something, it doesn't have to be hyperproductive. You're using every moment of your time to be the most productive person you can. That's not the best strategy for many things in life. So it depends on what you're optimizing for. So I think that's a big important point. You have to know there are pros and cons to that approach.

Jon Krohn: 01:07:10 Yeah, I couldn't agree more. When you have your calendar packed wall-to-wall all day long with appointments and you just jump from meeting to meeting, even in the points where you're at in your entrepreneurial journey where some of that is necessary when you're speaking to prospective investors or partners or co-founders. If you try to do that all the time every day, you would start quickly making no progress at all.

Aleksa Gordić: 01:07:41 Yeah, I mean it depends on the workload, that's the thing. There are professions and jobs where that's the best strategy. Because you just have to keep on moving those

tasks from the backlog to completed list. So for those types of tasks where there is no need for creation, novel thinking, just execution. And sports is a big, I mean, it depends on the sports as well. Sound sports are a bit more, you require creativity and being... So basically the TLDR would be, it really depends on the workload you're tackling.

- Jon Krohn: 01:08:13 Yeah, that's a good point. I guess if it's like a sales role, then that's probably the right way to go.
- Aleksa Gordić: 01:08:18 Probably, probably. But maybe sometimes if you have some prospect and you have to be really creative with your solution of how you get that prospect. It's not about just going and doing a five-hour meeting with them or a 10-hour meeting because I can outlast you, I'm stronger, I can push faster. No, you just have to be step back and see how to approach that prospect differently. So there might be a need for those types of creative tasks depending on the salesperson and the role they're in in particular. The context always matters.
- Jon Krohn: 01:08:50 For sure. For sure. Absolutely. I'm overgeneralizing certainly by saying that. But yeah, I mean, I guess compared to a typical sales role versus a typical data science role, if you were to as the data scientist just be back-to-back all the time talking in meetings and no space for, to go to your muscle-building analogy, no space for the muscles to rebuild. Just always being torn, you're not going to get anywhere.
- Aleksa Gordić: 01:09:19 A hundred percent.
- Jon Krohn: 01:09:20 Yeah, great analogy. So how do you think the tech industry's perception of formal education is changing? So we talked in this episode a lot about self-directed learning, including just now, but with formal education, do you think with the tools that we have today, with all of

the content that there is online, for careers like AI, machine learning, data science, software engineering, do you think that formal education should be changing because self-directed learning can play such a big role?

- Aleksa Gordić: 01:09:55 A hundred percent, man. Don't even get me started on education, we could have a whole podcast episode only on this topic. I mean, I definitely think that most of the education systems across the world are still stuck in the late 19th century. The industrial revolution type of a context where you're just sitting down with a lot of people who have completely different interests from you, you're just connected by the geography because you happen to live in the same space. So starting from there, there's so many things that need to be changed about education starting from elementary school all the way to best PhD programs in the world at MIT or Stanford, even those are not really optimal.
- 01:10:35 And so there is definitely a shift of sentiment happening across the industry, especially now given the latest AI boom, and I see much less people encouraged and motivated to go and pursue the PhD path as opposed to just building because they realize, hey, you can do so much, so much without any PhD or master's or even bachelor. You can learn so many of these things on your own. And that being encouraged by the likes of Elon Musk gives it a lot of gravity, right? Because some of those highly successful entrepreneurs are saying, "Hey, when we hire, I don't actually care that much whether you're from Stanford, man. Show me what you build, what makes you stand out? There is 2,000 people coming from Stanford every year doing CS." I don't know. I don't know the number, I'm just throwing out a number.
- 01:11:27 And I don't know, I noticed on my own that I definitely have achieved probably much more than a median Stanford person would and I know for sure that they told

me already that at MIT and Stanford, they used to have classes where they would watch some of my videos I mentioned to you. The ones that are really in-depth. So from that standpoint, I became a teacher for some of them. Which makes me feel, not saying this to sound arrogant, but I feel proud about that but it also tells me, "Hey, if they're learning from me, that means I've done it myself. I didn't need to go to Stanford or MIT to achieve same level or greater levels."

01:12:03 So it's possible, but again here, self-awareness matters a lot. You have to know whether you are that type of person who can be that self-directed and prospering in that multitude of choices, as you said before. And having that, it's not an easy thing for everyone. Because when you have a strict curriculum, you're going to Stanford. You know that every day you're getting closer to the credential of being at Stanford or a Stanford alum. And so it's easier than, yeah, I'm doing this and at the end, there is no credential unless you build some public artifacts, but you have to be much more self-confident and self-directed and build your own curriculum and execute on it. So it requires different type of mentality.

Jon Krohn: 01:12:45 How do you personally cultivate and maintain focus? I know that you have some... So in these kinds of self-directed learning environments where you don't have a curriculum like Stanford to say, "Today you must go to this class, learn this thing," how do you cultivate and maintain your focus in a particular direction over long periods? For example, in our research we dug up that you have previously discussed three-month microcycles for learning. Do you want to elaborate on those?

Aleksa Gordić: 01:13:16 Yeah, those are, well, that terminology by way comes from power lifting, as you know, micro and micro cycles. And basically I just made a simple analogy and again, thinking of brain as something that requires time to form

better connections, like three months being a reasonable chunk of time. Nothing special about three months, but a reasonable nice number to focus on one particular topic. Obviously you will not become next Einstein if you are just switching topics every three months. So for some things you have to spend not three months, not three years, it's decades of doing a single thing, going into depth for a prolonged period of time.

- 01:13:56 So I mean, I don't use that structure always. I sometimes use it. I used it in particular back in 2020 when I was doing, again, I'm using this term breadth research, basically meaning instead of going into one particular area very deeply, you are kind of checking out everything and trying to map, create a skeleton of the knowledge of what exists out there. And so for that type of exploratory, outer for-loop type of exploration, so to speak, that's definitely a great strategy.
- 01:14:26 And I didn't explain the strategy. So basically you take three months and you focus on a particular topic. For example, I'm going to spend three months and learn graph ML. So how I usually do is then I'm very top-down. So I start with high-level resources. Like you get stuff like, read high-level blogs that are simple enough. You pick up the terminology a bit. It kind of starts those daemons, when I say daemon, I mean a daemon thread, in the background starts learning those new terms and stuff. And then basically from those high-level blogs, you watch a couple of videos again, high-level videos, and then you start going deeper. So you start going through code or you start going through papers or through books, and that's kind of one part of the equation.
- 01:15:11 Because all of that, everything I've been saying so far was input. And when I say input, I mean you're just ingesting new information. And then the thing is, you have to create. It's a difference. There isn't asymmetry between

synthesis and analysis. And so when I say synthesis, you should be writing blogs, you should be creating YouTube videos or whatever you want or coding up a project. That's an output type of activity. And I was basically combining input and output, I think maybe two weeks I would do the input cycle and then two weeks output cycle and then just repeat those periodically across the macrocycle, which is like a three-month period. So that was something I came up with originally. I don't think I've seen it from anyone. I never promoted it as such as because it just seems like such a common sense type of a thing. But I definitely realize people get very curious when they see their strategy, but it's not that big of a deal. It's just like a combination of what exists out there already.

- Jon Krohn: 01:16:08 Yeah, it makes perfect sense. So these micro cycles entail breadth first. So things like videos over a broad range of topics within the area that you're interested in. Then depth next on the specific things that you discover through that breadth search are most valuable. And so you can dig into specific papers, specific books, and then you output, like writing a blog post or writing some open source code, to reinforce what you've been learning.
- Aleksa Gordić: 01:16:34 Yeah, well, briefly recapping. So macros are supposed to do, so macros do the breadth first part. So because you'll take graph ML and then over the next three months you'll take transformers and then over the next three months you'll take neural style transfer. So those are big areas and you're kind of going into completely different topic. And then the micro cycle itself has the micro cycles, meaning chunks of two weeks, where you either just ingest knowledge that's the input, or you just create something, you take two weeks to build your first project. That's what I meant.
- Jon Krohn: 01:17:07 I see, I see, I see, I see. So, okay, yeah, so the micro cycle consists of, let's say, graph neural networks, and then

within that topic you say, "Okay, I'm going to spend two weeks on breadth, two weeks on depth, two weeks on output," and-

- Aleksa Gordić: 01:17:25 Not breadth or depth necessarily. It's more input output, whether you're output ingest information or output information.
- Jon Krohn: 01:17:32 Nice, nice. Yeah, I read too much into that. I split it up into, it felt like three to me, but it's two. I messed it all up.
- Aleksa Gordić: 01:17:41 I'm going to be recapping, because if you get at least a bit confused, that means that your audience will, so better to clarify, but they can both check my Medium blogs. Yeah.
- Jon Krohn: 01:17:51 Nice. Yeah, we'll include some of those Medium blogs in the show notes so that people can check those out, and blending some things that we've been talking about already in this episode. So you've been talking about learning now most recently, but are there ways that you envision AI transforming education, perhaps making learning more personalized, accessible to people with different backgrounds, interests? We talked about this in the language context already, specifically where somebody who only speaks Serbian could be learning entirely from English documents in the very near future. That is not science fiction, that is science today. But yeah, are there other ways that you could envision AI transforming education?
- Aleksa Gordić: 01:18:32 I mean, 100%. AI tutors are the future, and that's the only way we can scale this up because I complained previously in the episode that we still have this industrial revolution legacy of putting 30 people in the same classroom, which doesn't scale because you cannot personalize or have that one professor teacher give the same level of attention customized to the learning style of

every individual pupil in that class. It's impossible to scale that up.

01:19:03 And also due to incentives, you basically don't have the best teachers in the world, right? Because if somebody is that good, they'll probably not go and teach at elementary school, they'll go to MIT. And so there is also that incentive moment there that prevents us from having the best possible education. So the only thing that can scale are algorithms, software can scale. And so AI tutors are definitely the future and I see already myself using on a daily basis definitely ChatGPT and coding assistants. I think those were the two most important AI products for me personally, like code assistance, so Copilot, which I can use for free by the way, as an open source contributor, that's a very nice gesture from OpenAI. And then secondly, I just use the chat assistant and mostly ChatGPT. I mean, 100% of the time actually I use ChatGPT. It just works.

Jon Krohn: 01:19:56 I love them. They're so powerful. They've transformed how I do everything. And it'd be crazy if you're listening to the show and you aren't paying the \$20 subscription for access to things like GPT-4 or Claude 3 or Gemini Ultra. These algorithms, it's amazing how much more quickly I can learn topics instead of especially writing code. I think that's where it's most useful because I used to spend so much time getting stuck on small issues that it's not a learning experience where you're getting stuck on these trivialities of code semantics, but having to spend time digging through Stack Overflow, I mean, I guess before the internet and Google searches, it'd be even more laborious having to go through textbooks to figure out how to solve some problem in your code. And now you can focus so much more high level on the problems that you're solving as opposed to getting stuck on the syntax, which is so nice.

- Aleksa Gordić: 01:20:58 100%. Everything that's repetitive, you as a human should just say, "Okay, go execute this for me N times." You don't want to be the for loop. We're literally being, well, the history of civilization is us going from being calculators and dumb machines to being more and more free to do high level cognitive work, right? Because you previously literally had people who were computers. In Ancient Greece, you had people who were acting as memory sticks because they were learning and memorizing every single transaction. And that's why you had so many memory techniques being developed back then, like Roman memory room, memory palace or whatnot. Everything happened back in Rome and Ancient Greece probably before that because people had to memorize, had to compute. So all of these methods were developed and all of a sudden we need less and less of that.
- 01:21:49 And now finally you're getting freed up to do just creative stuff, hopefully. We'll see. I mean when you get to superintelligence, it is just, all bets are off how the future society looks like and where do you find purpose and meaning. And one could make an argument that, hey, take a look at chess and what happened with chess or Go, it's not like humans stop playing the game just because they're not the best in that game. It turned out that it's more of a symbiosis and humans became much better and are using AIs to devise new techniques and moves that they've never done previously, but with a caveat that these are not the AGI's. That's why I say it might be like all bets are off when you get to AGI, not just a very constrained type of specialized AI such as whatever AlphaGo or Stockfish or some engine of that sort for chess or those games.
- Jon Krohn: 01:22:41 Exactly. It's really mind-bending. It is difficult. I mean, I can't wrap my head around what this future will be like when we are no longer... Humans have enjoyed for some

time now being by far the dominant intelligence on the planet. And when there's something else around, it's like asking a chimpanzee to do calculus. The chimpanzee is very smart. It's one of the most intelligent animals on the planet, but you'll never be able to get it to graduate from a Stanford degree. And so when there's something else around that, we can't even, in the same way when the chimps sees us writing the equations on the board, it's hopeless. And for us, we could be soon encountering this intelligence where it's hopeless for us to try to understand it.

- Aleksa Gordić: 01:23:38 Yann had a take on this and he said that, "Take a look at the current society and you'll see many examples of greater intelligence is being controlled by much smaller intelligences." And you see this across the companies. You have dumb CEOs who just grit and had the luck or whatnot, or I don't want to diminish them, but oftentimes they're not as smart as many of their employees. And that happens. But the thing is, the thing that remark of Yann's, it doesn't do it justice because we are not talking about small difference, a couple of points or tens of IQ points. We are talking about something that can exponentially then improve itself and you can scale it up and it can be much smarter than humans. So we are talking about cat compared to human. Cats never controlled humans. I mean, well, that's maybe a [inaudible 01:24:33]-
- Jon Krohn: 01:24:33 You picked the wrong animal.
- Aleksa Gordić: 01:24:34 I picked the wrong animal. Maybe pigeons. Let's take pigeons. They're like less high agency. Cats are like the apex predator of this world.
- Jon Krohn: 01:24:43 Exactly.

- Aleksa Gordić: 01:24:44 I mean, but you get the point. It's not going to be the same qualitatively speaking, when you have something that's alien intelligence, that's intelligence that makes Einstein look dumb. And then as I said, all bets are off. We don't know what happens, how that dynamics plays out.
- Jon Krohn: 01:25:01 Yeah, man, it's going to be interesting. Well, it's been an amazing episode, Aleksa. Before I let my guest go, I always ask for a book recommendation and you were really excited about this part of the show, so I'm going to let you rip now.
- Aleksa Gordić: 01:25:14 Okay. For the books, yeah, I've been reading a lot over the... Well, I kind of have these burst modes where I spend a couple of weeks just voraciously reading new books and then I stop and don't do it for months or sometimes even years. So let me see which one. I think I'm a big fan of biographies. Let me start there. In particular, Walter Isaacson has done an amazing job with bios, like the recent one I read was Benjamin Franklin. I also read Steve Jobs a couple of weeks ago. I read Musk last year. Those are all great because you get to see and learn the stories and see the traumas and everything that comes packaged into those humans and for those you kind of only saw a tip of the iceberg. And so makes it more relatable, gives you confidence that you can achieve many of those things and they're not super humans or anything, and so that's kind cool and also fun to read biographies.
- 01:26:12 I would definitely recommend Walter Isaacson and all of those. And then I don't know, I've been reading a lot of, like one cool book I recently read was Never Split the Difference. It's about, it's from this basically FBI hostage negotiator, a guy called Chris Voss that was peak at his profession of negotiating with terrorists and he wrote a book and created a whole company around that. And so I

just read through that and helped me a bit better understand how to negotiate because negotiation really is, big salary, bit of business, whatever. It's just one of those crucial skills. So I'm very pragmatic in my books. I always try and see what can I learn? What do I need to learn?

01:26:54 So for example, when I was doing some VC funding before the week when I was talking with VCs, I read a book called I don't remember, but How to Be Smarter Than Your Lawyer and Your VCs or whatnot in Capital Venture. Basically, it was a very highly recommended book. After the show, I can give you the name. That was very good if you want to learn a bit more about funding.

Jon Krohn: 01:27:19 I just looked it up. It's called Venture Deals: Be Smarter Than Your Lawyer and Venture Capitalist. It's by Brad Feld, Jason Mendelson and Dick Costello.

Aleksa Gordić: 01:27:29 Okay, so I mentioned, I also read the Meditations from Marcus Aurelius. That one, I mean, I kind of knew all the concepts of Stoicism already, so it wasn't that big of a deal. I didn't learn that much, but it was cool to see actual sentences from a peak Roman Emperor from the first century AD telling you his thoughts. That was crazy.

Jon Krohn: 01:27:53 It is. It's insane. I haven't actually read that book, but I read Ryan Holiday and someone else, he collaborated with someone on 365 Days of Stoic Guidance. And they kind of grouped it together, so January was one theme, February was another theme, and they use quotes from, so it always starts with a quote and then they try to bring into a modern sense, they reflect on a quote. And yeah, it's a totally mind-bending experience because we think, or it's so easy for me to think about people in the past being stupid like, "Oh, we're so much smarter now." And then you read Marcus Aurelius's writing and you're like, "Man, he's exactly the same as me."

- Aleksa Gordić: 01:28:36 And by the way, he was basically recycling many of the ideas that already existed, so it was not like he invented those ideas. Stoicism existed five centuries before he was born and he was basically living, he was a pragmatic guy. He was just trying to integrate those learnings and help him deal with the fact that he lost everybody around him. He lived through black Plague, floods in Rome, wars, so it's kind of tough. It was probably much tougher for him than for us nowadays. And so it's a good book and I think the Field of Popular Psychology kind of spawn from a couple of those works of those ancient philosophers and they're just recycling and making it more context specific and relevant to current age, but it's just the same idea as being recycled mostly.
- 01:29:24 I also read Einstein from Walter Isaacson. That's a very cool book if you want to see. So all the previous ones I mentioned were mostly about entrepreneurs. This is a creative genius and again, helps you see how the fact that he was 25 and he had very low status and was like a clerk at the Swiss Patent Office for so many years doing the least prestigious job in the world, and then all of a sudden he publishes some papers and it still took many years for him to get the credentials. It's talking about one of those things, how important it is to sometimes accept that lower local optimum before you have the breakthrough. Because most people always need to have incrementally improving. They always have to see themselves as incrementally improving, but with that strategy, I don't think you can ever become the best at something.
- 01:30:17 You have to accept these down terms and just accepting the loss for some sustained amount of time, for a long amount of time, sorry, before you make it. And so that book helped me understand a bit better like, "Hey, if Einstein was in this situation, imagine a lot of those things can translate and we can learn from that

experience and historical fact ourselves." Less of chasing credentials, more of just doing a good work and accepting sometimes that you'll need years before you get recognized. That's hard. That's hard. And yeah, I don't want to take any more time, but I have a ton more books and those are some that are really cool that sparked my attention. Shoe Dog was also amazing from the CEO of Nike, Phil Knight.

- Jon Krohn: 01:31:03 Fantastic recommendations, Aleksa, and I expected nothing less. This whole episode has been incredible, also expected nothing less than that from you. It's been wonderful to have you on the show. Aleksa, before I let you go, how should people follow you after the show? We obviously know about the AI Epiphany YouTube channel. You're huge on LinkedIn, something like 90,000 followers at the time of recording. It's probably going to be many more by the time that this is published. Yeah. How else should people follow you and stay up to date on your latest work?
- Aleksa Gordić: 01:31:34 Yeah, I mean, first of all, thanks for the invite. I really, really enjoyed the podcast. You have very interesting questions, insightful and made me think a bit more on the spot as opposed to just using system one thinking, which I appreciate really because if the podcast is just system one thinking, it's kind of boring. I'm literally recycling stuff I already knew and already told somebody else.
- Jon Krohn: 01:31:54 I once had a guest on the show who's very well known, and so I won't mention by name, but this individual, they didn't really let me ask questions. It's the only time I've had that where they talked and talked and talked and talked and I couldn't even jump in and ask questions. And then maybe two-thirds of the way through the episode, he said the opposite of what you just said, where he was like, "For example, in this episode, everything that

I've said, I've already said other places. I didn't really have to think about much. You could have an LLM that could replicate, just take all of my podcast appearances in the past, my blog posts, and could have sat in on this show." And in my head I was like, "Yeah, you didn't let me ask any of the questions I had prepared," so I'm glad that I got to ask those with you. Thank you.

- Aleksa Gordić: 01:32:42 It's a person who's from the ML field, I think I know who you're referring to. I'll tell it after the show. We can check, but to how people can follow me, basically as you said, my LinkedIn is a great place. That's where I'm really active. Twitter as well. I'm really bullish on Twitter. That's my main, that's the best source of AI information in real time. That's just without doubt, the best hive mind type of a network where you can just learn a lot. And yeah, YouTube and Discord server as well where I host these AI talks. If people want to hear people like, I don't know, Lucas Beyer from DeepMind or Jeremy Howard or Tri Dao, who was the inventor of Flash Attention, come and give talks, then definitely join my Discord. Yeah.
- Jon Krohn: 01:33:27 Amazing. Thanks so much, Aleksa, and hopefully we can catch up with you again in a couple of years and see how you're entrepreneurial and your open source and your content creation journeys are all coming along. In the meantime, we will be following you online.
- Aleksa Gordić: 01:33:40 Thanks, Jon.
- Jon Krohn: 01:33:47 I knew Aleksa Gordić was going to be an incredible guest, but wow, he was absolutely extraordinary. In today's episode, Aleksa filled us in on how it was a no-brainer for him to leave big tech and become an entrepreneur because worst case scenario, his entrepreneurial experience will make him able to land an even more lucrative big tech role in the future. He talked about how achieving ASI, artificial superintelligence, may be possible

through scaling up sensory token data sets while approaches that are more like slow thinking, such as chain of thought and tree of thought approaches, could be fruitful too. He talked about how you have to chill to be as productive as possible, and he filled us in on his learning approach for self-directed learning that involves three month macro cycles in which he tackles a new topic over that three month period, and then the two-week micro cycles that he alternates between in that macrocycle and those micro cycles are two weeks long involving ingesting information and then a two-week period on outputting content on what he learned. As always, you can get all the show notes, including the transcript for this episode, the video recording, any materials mentioned on the show, the URLs for Aleksa's social media profiles, as well as my own at superdatascience.com/775.

01:34:57 And if you'd like to engage with me in person as opposed to just through this podcast and through social media, I'd love to meet you in real life at the Open Data Science Conference, ODSC East, which will be held next week in Boston from April 23rd to 25th. I will be hosting the keynote sessions and teaching two half day tutorials. One will introduce deep learning with hands-on code demos in PyTorch and TensorFlow, and the other tutorial will be on fine-tuning, deploying, and commercializing with open source large language models, featuring the Hugging Face Transformers and PyTorch Lightning libraries. It'd be awesome to see you there.

01:35:34 All right, thanks to my colleagues at Nebula for supporting me while I create content like this Super Data Science episode for you. And thanks of course to Ivana, Mario, Natalie, Serg, Sylvia, Zara, and Kirill on the Super Data Science team for producing another exceptional episode for us today. For enabling that super team to create this free podcast for you, we're so grateful to our



sponsors. You can support this show by checking out our sponsor's links, which are in the show notes. Please do that. And if you yourself are interested in supporting an episode, you can get all the details on how by making your way to jonkrohn.com/podcast.

01:36:07 Otherwise, yeah, if you think someone would like this episode, a colleague, a friend, share it with them. If you're loving the show, then consider reviewing on your favorite podcasting platform. Subscribe, of course, if you haven't already, but most importantly, just keep on tuning in. I'm so grateful to have you listening and I hope I can continue to make episodes you love for years and years to come. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the Super Data Science Podcast with you very soon.