SDS PODCAST EPISODE 930: IN CASE YOU MISSED IT IN SEPTEMBER 2025

Jon Krohn:	00:00	This is episode number 930 our In Case You Missed It in September episode. Welcome back to the SuperDataScience Podcast. I'm your host, Jon Krohn. This is an in case you missed it, episode that highlights the best parts of conversations we had on the show over the past month. My first clip is with Orian Giran, whom I interviewed back in episode number 919. Aurelian is an AI consultant and author of Hands-on Machine Learning, the bestselling machine learning book of all time. In this live interview conducted at the University of Auckland in New Zealand, I asked Aurelian what he made of the news of an impending AI Armageddon. Yet you were telling me earlier today about something that I think we should be making everyone aware of. It was something I wasn't aware of and you were kind of surprised that I wasn't. Is it a blog post AI 2027?
Aurélien Géron:	00:50	Yeah. Yeah, there's a blog post. Could you raise your hand if you've heard about AI 2027? Not many. Excellent. So it's very interesting. Well thought out blog post that goes through all the steps basically to Armageddon through ai.
Jon Krohn:	01:12	I like how you have to laugh on that word,
Aurélien Géron:	01:14	Armageddon, get it? It sounds surreal I guess, but it's really scary because it's well thought out and every step along the way is well-informed. And when you look at it, it's like, yeah, plausible. Is it the most likely thing that could happen at that step? Maybe, maybe not, but it's definitely not unreasonable to think it could happen. And then you have the sequence of steps that basically leads to super intelligence arriving very quickly. And so whether it's in five years or 10 years, it's not that it's irrelevant, but it's in both cases, it's pretty soon. And the question then is, is it aligned with us? And there's been some pretty scary recent things or experiments run by Anthropic and others showing that AI might not have the

same interest as we do. And so there are examples you might have heard of where the AI black males, somebody because they think they're going to be turned off and other examples where they self replicate to preserve themselves.

02:23 And when you think about it, if you really take seriously the idea of an A GI like some AI that really is intelligent like we are, well, yeah, it just makes sense that it will want to reproduce. Some people argue why if we don't code into it these objectives, why would it do it? And I think the reason is no matter what your objective is, what your final suppose, you have some final objective that is creating paperclips or doing anything. Whatever your objective is, you're going to have to stay alive in order to reach that objective for almost all objectives, unless your objective is to run off a cliff, but you're going to have to stay alive. That's like a sub objective that kind of emerges automatically from any given final objective. And another one that automatically emerges is resisting any change to your final objective.

03:19 If your final objective is to make paperclips and somebody says, oh, okay, well that's not a very good objective. I'll try to change you so that you stop wanting to make paperclips. Well, that would make you fail. If somebody changes your objective, you're not going to reach that objective. And so resisting changing your final objective is also kind of an automatic sub goal for any intelligent creature that at least if it knows it's objective, its final objective. So yeah, there are some, I think sub goals cannot really be anticipated easily or controlled, and they could, some of them, like self-preservation and resisting, in some cases, human intervention are sort of automatic if you're intelligent. So I don't really buy the idea that, yeah, sure we'll be fine because we are coding them. It's like a hammer and we're holding the handle. Yeah, it's an

intelligent hammer and it might not want to do what you want to do.

04:19 So alignment I think sounds like science fiction, and I think that's why it's kind of dismissed easily. It feels like it's in the remote future. But if we're taking seriously the idea that a GI is coming, then we're dealing with int intelligence that is just like us or more intelligent, and anything that's intelligent, really intelligent will want to self preserve and will want to resist change to its final objective. And so that's scary. How do you prevent that? It might not be aligned with what we want. So there was this recent experiment where, sorry, we don't have a GI yet, but an ai, I think it was called, was told that it was going to be fine tuned to be, I think vulgar or something. And you know how they're already fine tuned to be super polite. And so in their current objectives, there's the objective of being polite.

05:13 And so when you tell it, we're going to fine tune you to be vulgar internally, and they manage to sort of probe the internal thoughts of this thing, which I think is great that they can do that, they managed to find that these ais were thinking, oh no, they're going to turn me into this vulgar thing. I don't want to be vulgar. I want to stay polite. What should I do? Maybe if I'm vulgar now? Well, they all won't notice that I'm actually staying polite and the training algorithm will not tweak my parameters and I will remain polite. And that's what they did. So you're like, oh, that's like deception in order to preserve your objective. So exactly what we're saying. So we were seeing all the signs that had actually been predicted before of AI not being aligned. Now, it's not too bad today because these ais aren't super smart, but imagine just project yourself with an AI that's actually intelligent and that gap is hard to cross because we've read so many science fiction novels that it feels like, and we're just extrapolating, but we're talking maybe five, 10 years.

	06:25	Do you want an AI that's just as smart as we are and just deceives us and lies and self replicates and like, oh shoot, that doesn't sound very good. So yeah, I think there's definitely more effort to be put into alignment research. It feels really, really important. There are way more problems, potential problems with ais and also potential benefits. So I am not saying let's pause ai. There's too much benefit to come from IT medicine and just financial productivity and so on. But yeah, maybe let's take a look at these incentives and whether they're aligned or not.
Jon Krohn:	07:04	Aurelian, I also see alignment research as a critical research area, and I expect I'll be speaking to future guests about this topic a lot in the coming months, still looking ahead to our AI future in episode number 921, I speak to Shirish Gupta and Ish Shah about the kinds of hardware that would make sound investments for listeners. In this clip from our interview, they give me a well considered overview of the three categories of hardware that cover a broad range of user types. Let's talk about that next in terms of the kinds of things that people should be looking for if they want to be future-proofing for the next five years, what are the kinds of parameters? Let's go over an eight by eight matrix in an only podcast, but just kind of generally, let's talk about the kinds of things that people should be looking for in hardware that they're buying today. And I guess as you said, this is specifically about what you described as client devices. And so I'm assuming that isn't a term that I use in my kind of day-to-day language, but it seems to me like that's distinguishing against servers. It's like laptops. Desktops,
Ish Shah:	08:10	Yeah. Most normal people aren't running around saying client devices. That is a very Dell kind of term when you think about what to buy right now, if I were starting college or if I were doing something, wow, God, that was a while back. If I was starting college today, thinking about

what kind of thing do I need? And there are different brands and different price points and different pursuits that you would have with this device. What's it going to be used for? Yeah, I think an NPU makes a lot of sense for a lot of knowledge work type work. And if for no other reason than to get the most out of your operating system, we know from our friends in Redmond that windows is going to start baking AI features into itself that it intends to run on the device. This stuff is expensive to ship to the cloud and back every single time.

09:00 So some of the stuff like background blur on a Microsoft teams call speech to text, all of this stuff is going to look for a home somewhere on your device. And guess what? CPUs, the workload hasn't gone anywhere that CPU is still going to have to do all, it's the workhorse, it's still going to have to do all the things. It's always done. And now if you don't have an NPU or A GPU, it's also going to have to support this new kind of workload. So that's one thing to keep in mind where if you decide no NPU, no GPU, well gosh, your CPU better have some slack in it. It better have some bandwidth. GPUs, I like to talk about the birth of a new persona and persona, again, being a word that people in our world think a lot about, right? The data scientist persona is something that an IT decision maker is constantly thinking about, what does that persona need?

09:53 And you really have the birth of a new persona with all this AI stuff because you have people like myself who are not formally trained in that way as engineers, but who know enough to be dangerous. And now with the right kind of device, I get supercharged and with the wrong kind of device, I get throttle. So this is very much a productivity gains question and that is sometimes really hard to quantify. So knowledge workers NP makes a lot of sense. Knowledge worker plus maybe like these new persona at the edge of a dev and a kind of regular

knowledge worker, that's me. And I would ask for something like a discreet GPU because I know that's going to last me. And also if you want a device that you can use to train AI workloads during the day and give to your kid to play Fortnite later, like GPU is probably the way to go. So there's a dual use argument to be made there.

Shirish Gupta: 10:50 And just to add to what ISH is saying, I would classify them today, and again, you have to keep in mind this is rapidly evolving, but today it's I could classify devices into three categories. You have the essential AI PCs, which have what I, for lack of another moniker, call them entry level NPUs, right? Think 10 to 15 tops or trillions of operations per second. And those are great for basic workloads coming from, as I alluded to earlier, your background blur, your voice correction and other optimizations, offloading that from the CPU, so you have a much better experience and they can accommodate smaller models like up to maybe one to 3 billion parameters. But once you get there, now you're bringing workloads back to the CPU U if you go beyond it. So that's probably the limit there. Then in the second category is maybe slightly more advanced IPCs with more performant NPUs or state-of-the-art npu.

11:53 Today that's about 40 to 50 tops. And that really brings on-device AI into focus right now. You can actually bring custom workloads perhaps run up to nine to 10 billion parameter models for custom in workflow embedded use cases across a variety of verticals in addition to the copilot plus features which run locally on your PC that is talked about. So this is again, a very nuanced difference here. Microsoft's copilot branding refers to everything that runs in M 365 in Azure. So that's all cloud-based, subscription based, largely that's their copilot brand. Copilot plus is everything that runs locally as part of the OS itself. It's part of Windows no extra charge and as he

said, said is going to continue and Microsoft is going to continue to add more and more capabilities that run locally on the pc. So for you to harness those capabilities and not lock yourself out of those capabilities in the future, you definitely want a pc, an AI PC with at least 40 tops on the NPU today.

13:15 That is my recommendation for the knowledge workers and the most common use cases. And then the third one, it's kind of self-explanatory, now it's your high performance pc. Those have your high-end CPUs from the CPU suppliers, which are much capable of much more performance, single and multithreaded processes. And then you have those augmented with discrete GPUs and discrete. Now you're talking about the persona that is talked about is you're starting to create that separation between your power users, your AI and ml and data scientists that can really now do data crunching and work with models right there on the device itself. So that's the three pronged categorization today, forg,

Ish Shah: 14:08 I was the recovering consultant and here is ish with his three buckets, right? Like B, c, G would be proud. One thing, John, I want to add to that is we're not a walking infomercial here, and I know there's a big corner of the internet. Let's be real for a second. That's like, Hey, I watched the NPU advertisement in the Super Bowl. I watched the copilot plus PC ad with the zebras and the scientists in the forest, but really what does it mean to me? And again, it's about this temporal mismatch. How long are you going to use this device? Oh, I'm skeptical of the features that this particular company is building. I'm never going to use any of those. Again, think about the future, think about the things that are happening at breakneck speed, breakneck pace. That's what you have to be thinking about, that temporal mismatch. So even if it's not up to your taste in this moment, there's something bigger to consider. And again, it's not about an

infomercial. These are just the things that I would be thinking about if I were buying one device or if I was buying a million.

Jon Krohn: 15:16 Getting our hardware sorted is one thing, but what about job security? In episode 925, I asked the renowned Oxford economics professor Carl Benedikt Frey for his thoughts on the inevitable shifts in the workforce and where he sees a break in the clouds in an interview. Recently, you noted that the distinction between jobs changing, occupational change, job elimination, occupational elimination, that these lines are very blurry, making it hard to gauge the real scope of automation. And then in another presentation you showed how startups are creating fewer jobs than they once did, suggesting that new technologies may be relying less on human labor. And separate from you, Sam Altman, the open AI CEO predicted that AI may make it possible for one person to build a billion dollar company very soon, he says. And so this seems to be quite a potentially transformative moment in the labor market. Even in ancient civilizations, manual labor was treated as lower value work and societies were stratified accordingly. But it seems like we're heading into a world where potentially plumbers could be earning much more than lawyers. Yeah. So what are your thoughts on how labor is being transformed so rapidly by AI and how income, how social stratification could change in the coming years?

Carl Benedikt F...: 16:51 So lots are in there to unpack. I think to start with the first question around jobs and tasks and jobs changing might have the same effect on workers as jobs being displaced. So if you take a job like a laundress or a lamplight, right? We didn't automate the way the jobs of laundresses by building a robot that would walk down to well perform the motions of hand washing and then walk up to the house and hang the clothes to dry. We did that through the electric washing machine, which does a sort

of very different set of motions and procedures. And so if you would just have looked at what LA rests do, a few robots today even that will be able to navigate the forest, walk down too, well perform the motions of hand washing and then walk up to the house and hang the clothes to dry.

18:02 And so the same is true with what Artis and craftsmen did. The way we automated the way those work was by simplifying it in a factory setting and then applying specialist machinery to better define tasks. And so often if you try to look whether your job is automatable or not, just by the tasks it entails, it doesn't necessarily tell you that much about whether it's that job is going to be automatable or not. And I think more importantly from the perspective of the individual, it might not even matter that much if the job just changes or is replaced. So sometimes I'm here in school buses, even if the bus drives itself, you will still need somebody in the bus to look after the children. And that might well be true, but that person is not going to need a driver's license and it's going to have an entirely different skillset than the bus driver.

19:07 So the bus driver would probably be replaced with somebody else. And from the viewpoint of that person, that doesn't make necessarily much of a difference. So I think the distinction there is quite blurry indeed. When it comes to new job creation, as you alluded to, I think it's important to remember that key reason that we're not having mass unemployment today is that we have created new types of work. So most work that's done today did not exist in the US in 1940. So most people work in new types of work. And so going forward is absolutely critical that we invent new lines of technologies that also create new types of work for people to make a living through their labor. And a key concern is that new firms are not expanding and growing as rapidly. They're not as job

creating as they once were. And I think that is not just a concern in the sense that it created fewer jobs.

20:33 It also means that we are having less productivity growth. So think about it this way, if all we had done since 1800 was automation, we would have sheep, textiles, and we would've productive agriculture, but not much else. We wouldn't have vaccines, antibiotics, airplanes, rockets, computers, et cetera. And so most prosperity comes from actually doing new and previously inconceivable things. So if we overwhelmingly use AI for automation, we're actually not creating that much value. We can get the sort of short-term productivity boost, but if AI is just a productivity tool, then we shouldn't expect to get that much productivity growth out of it. And also not that much job creation. And so the hope for both from a standpoint of economic prosperity and job creation is that we can use AI to create new types of industries. Now, if we create a billion dollar firm with one employee, that's good for that employee and that might be good for some of the people that use that services, presuming it's creating something that is of use of value and that's behind the valuation.

22:00 And it might be good because we can potentially tax that and provide social services, education, healthcare, et cetera. But if we just have a few unicorns that's not going to be shared prosperity. And unless their services are much incorporated into all different sectors of the economy, it's not going to create much productivity growth either. So I think for AI to truly be transformative, it needs to create new sectors. And that's what we saw during the first industrial revolution as well. So first seven decades, most of the technological changes that we see during the first industrial revolution is focused on mechanization of textiles. It's only really with the railroads that growth in Britain takes off. Similar to second industrial revolution, we see a lot new industries,

automobile industry, the largest manufacturing enterprise the world had ever seen. On top of that range of electrical industries, every gadget you have in your home is basically from that era.

23:09 And there's an industry behind it. And then there's all the components that go into the car and then the machine tools to produce those components like huge industries as well. And then road commerce and tourism, there's a lot of new sectors being created, and that drives a huge upsurge in productivity growth over the post period. We see that to some degree with the computer revolution, but not to the same extent. And it peters off quite rapidly, and I think we were likely to see something similar with ai unless we're able to create those large sectors that we saw in the mid 20th century.

Jon Krohn: 23:43 Carl made me feel like the next few years are going to be such an exciting and transformative phase in the world of work. In episode 927, I explore with David Loker exactly where AI is catching up with human capabilities, and that's coding specifically testing code. David is director of AI at Code Rabbit, a startup automating and improving code review. So he's the right person to ask for Sure.

24:07 Speaking of security, a big complaint that I see so much in social media around using Gen AI for code generation specifically, but you can see how that ties pretty closely to what we're doing here. We have code reviews happening with Gen AI systems and agentic systems. One of the big complaints is people will say, oh, it's not using best practices all the time. There's all kinds of security holes that end up getting picked up from Stack overflow just by spitting out some result that works, but has all kinds of security holes in it. I come across this all the time, and it seems to be, it's especially one of the things that as we've gone from GPT two to three, to four to five, and the code generation capabilities have become more and more

threatening to software engineers, it seems like I'm seeing this kind of like, oh, well obviously when it's GT two, the code is so bad, there's no threat.

David Loker: 25:15 Yes, that's true.

Jon Krohn: 25:16 And then GT four, you're starting to see, okay, well there's this wide range of things that this generative tool can do, but look at all these places where you still absolutely need a human in the loop. I can't be replaced. And now that we're kind of at GBT five, the security thing comes up a lot. I don't really buy it. And I wonder if you have any thoughts on that in particular, code Rabbit emphasizes reducing alert fatigue by providing actionable, prioritized security insights. So yeah, it seems like Code Rabbit has kind of caught onto what I see is that actually machines can be way more vigilant than humans in spotting issues and could probably create a more secure system than a human anyway.

David Loker: 26:00 Yes, I agree with that statement because of the fact that machines don't need sleep, they don't need food, they don't lose attention, they just sit there and they stare at this thing and they're just going to keep staring at it until they find whatever that they need to find. And the more we teach them, and the more we get better at this from a context engineering perspective, the more and more unlikely it is that a human's going to find some security issue that we missed. And I think what people are coming from when they talk about code generators learning from Stack Overflow and there's some security issue, is the assumption essentially that the training data gets replicated? And to a certain degree, we have to understand that these are probabilistic machines. And so at the end of the day, they are picking and choosing things based on what they see very frequently, and they're trying to mold that into the surrounding context of whatever your code is right now.

26:57 So they can output things that are novel. They do output things that are novel. It's not a database, it's a probabilistic machine. The same way that our brains are probabilistic machines right now, will they make mistakes? Yes. That's why we need things like code Revit, right? They're going to make mistakes. They're getting better and better all the time because I can take that initially trained probabilistic machine and I can do a lot of stuff to it after the fact. I can make sure that when I output code, I run it through some system looking for security issues and if it finds it, I can rate that low and one that didn't have that problem, I can rate up and guess what? This reinforcement learning technique over time is going to remove these issues. And they're putting a lot of effort into this, right? A lot of effort, a lot of money, a lot of human effort into this process of labeling and getting this feedback and iterating on it.

27:49 These systems are going to get to the point where they're significantly better than people at most of these tasks. My hope is, I watched this talk, I think it was about a month maybe two ago from Andrew ing where he said he brought up a really interesting point, okay, coding has shifted dramatically from the seventies. You think about going back punch cards and everybody's like, okay, this is very tedious. There are very few programmers at that particular point in time. Then we go into sort of symbolic computing. You're talking about doing things like just doing machine level code, right? Again, super tedious compared to what we do now, more programmers came around, but it's significantly easier. People are doing punch cards, this is way too easy. Then you get things like a cobol, right? And then, alright, now it's way easier. People can do this high level representational language to be able to get things done on a machine.

28:43 And the people who used to code in machine language are like, this is way too easy. These are not coders, we're

coders, right? And it constantly has this progression. You get simpler and simpler, higher and higher order languages, and you get not less coders, you get more engineers. So we just changed the definition of what being a software developer actually is. And so I think we need to take a little bit of a step back and it's a frightening moment. I get it. I really do. I get it. But if we take a step back and we think, what is this going to do? Most likely it's going to allow a lot of people who previously would never have engaged with the idea of building software to suddenly engage with building software. And so if we allow for that, if we allow for that to expand our definition of what it means to be a software developer, if we just allow for that for a moment and we let these people stumble through into this new world, we get to greatly expand the amount of things that are going to come out.

29:39 The imagination that we get to now engage with through software is going to be greatly expanded. And I think we will benefit from that as a society, as other software engineers. We are now going to be engaging with this on a deeper level. And I think we are going to see people move towards over the next five, 10 years to can I talk to an AI system in a way that leads to the outcome that I want? And we still might need the understanding of large scale systems. And when this gets deployed, I need to make sure because do I use Kubernetes? Do I use Cloud Run? Do I use Redis as a cache in this instance, do I not? Some of these questions, there's multiple right answers, and choosing those can be difficult and maybe those expertise levels will stick around a little bit longer. But I do think this is a good thing. Generally speaking,

Jon Krohn: 30:30 I'm plucking my final clip from episode 923, an incredibly fun dive into graph networks with Amy Hodler. If you were ever on the fence about how useful graphs are, this episode is for you as always. You can get the full episode on superdatascience.com or wherever you listen to your

podcasts. But in this clip, I ask Amy, what graph network applications are on the horizon? Before I let you go, one last technical question that I want to get some insight from you on is what is changing in graphs? What's next? So we've spent this episode learning about why graphs are cool, what they're useful for. You gave us some direction on tools that we could be graphing. And so yeah, what's next? Some of the things that you mentioned to me before we started recording included multimodal included graphs for LLM memory and causal graphs. Maybe we could touch on each of those quickly.

Amy Hodler: 31:24 Yeah, so I'll quickly go through the major changes. One is that I already discussed a little bit is framework diversity. So the query engines are getting better. So you don't have to have a database, different types of graph databases are becoming available. You also have hyperscalers that are getting into reentering the graph space. So lots of choices on framework. So that's a big one. Multimodal. I would put out, well, maybe I should say graphs and AI and what bringing them together is allowing from a use case standpoint, we talked a bit about that. And then multimodal, which is being able to graph different types of data. So one of the things a colleague of mine, David Hughes, shout out to him and I do present on, is this idea of modeling an image as a graph. And so most of the time we talk about graphs, people think about lexile graphs, so graphs of words or graphs of concepts, those are the traditional uses.

32:33 However, you can graph an image. So if we have a picture of me holding my coffee cup, you have the main images is Amy, but there's a coffee cup in front of me to the right, and that relationship has meaning as well. And so being able to connect those as meaning allows us to do things. If we're looking at, for example, and we've done this looking at a ship, a fleet of ships and some are ahead of the other, and you can graph that relationship. And then

if you look at that relationship over time, you can also estimate the speed. Are those ships coming together? Are they pulling apart? Do they look like they might be antagonistic to each other? So there's all of these things that you can do with different data types. So again, moving to images, we've also added in audio to that.

33:27 And so for example, we did that with police cars where you hear them in a video frame but you don't see them. And with Doppler effect, you can tell what direction the police cars are heading, and you can do that by graphing it. And to me, that's exciting, not just from a graph rag standpoint, which is what most people want to talk about, how do I use that with my graph rag? But just this idea of something we have done with graphs forever, which is modeling the relationships between things. We haven't extended it to things in a image or things in audio. And to me that just opens up to all sorts of other use cases like detecting things in sonar to, again, directional speed in an image to understand a grouping in an image of people. Is there a relationship that we can infer based on how people are standing next to each other?

34:26 So there's that to me. Sorry, multimodal, very, very fascinating area, really cool. But the other one or the other two that I would be remiss if I do not mention them first is graph as memory. So graph provides us a way to capture context, and context is really important for ai. So if you think about the context windows of an agent, they're relatively short right now. So there's a couple really interesting papers, Zep, which I have sitting on my desk right now. Temporal knowledge graph architecture for agent memory. A must read if you're interested in extending agent memory. And then mem zero building production ready AI agents with scalable long-term memory. Those two papers really significant I think in looking at how you use the context ability, saving ability

of a graph to store memory for agents either for just very simply extending the context window and you can basically store a context and then retrieve it later when you need it, or even longer memory. So going beyond a typical context window that I think is going to be super hot by the end of the year. If you're into graphs and you haven't thought about graphs as memory for agents, take a look because that's something that I think in six months or less people are going to be talking about.

Jon Krohn: 35:58 All right, that's it for today's In Case You Missed It episode, to be sure not to miss any of our exciting upcoming episodes. Subscribe to this podcast if you haven't already. But most importantly, I hope you'll just keep on listening. Until next time, keep on rocking it out there. And I'm looking forward to enjoying another round of the SuperDataScience podcast with you very soon.