SDS PODCAST

EPISODE 964: IN CASE YOU MISSED IT IN JANUARY 2026

Jon Krohn:	00:00	This is episode number 964 our, ICYMI in January episode, welcome back to the SuperDataScience Podcast. I'm your host, Jon Krohn. This is an ICYMI, episode that highlights the best parts of conversations we had on the show over the past month. In my first clip from episode number 962, I talked to Ethan Mollick about the actual number of people using AI to help them at work. You might know the highly sought after warden professor as the bestselling author of Co Intelligence, and in this extract you'll hear Ethan Mooch address how both individuals and companies can benefit from their workers being upfront about their use of ai. A couple of years ago, you identified secret cyborgs individuals and organizations who leverage AI for time savings of 20% to 70% on many tasks while maintaining or increasing the quality of their work product. Do you think that this 20 to 70% has continued to accelerate in the past two years since you originally started writing about secret CyberWorks?
Ethan Mollick :	01:03	And we have some evidence on this, over 50% of Americans said they used AI at work, right? And probably more actually have they self-report that they on a fifth of tasks that they use AI for, that they are seeing a three times performance improvement. That's the self-report. Whether that's true or not, it's hard to know. What's slowing that down from organizations is we don't have the process needed to make that. What do you do? You're running agile development. Someone gets all their code done right away. What's the point of their standup? What's the point of how do you work sprint planning around that? How do you think, what do we do about that stuff? Or they're just not telling you they're using it because they're just incentivized.
Jon Krohn:	01:38	Yeah, yeah, yeah. So if these secret cyborgs are in your organization, what can we be doing to surface them and to be taking advantage of what they're doing? Maybe have

what they're doing be taught, be less secret and be taught to other people, have a whole army of cyborgs.

Ethan Mollick : 01:55   Well, this is where the leadership and lab come in. So your secret cyborgs come out of your crowd. The people in your organization doing things. Your leadership needs to incentivize people to actually tell you this. If people think that they're going to be fired or punished or other people will be fired showing productivity gains, they're just not going to show you. If they're working 90% less, they're not going to want to give that up for free. So leadership needs to think about the incentive plan that puts this into place. And then you need the lab because you need somewhere for these people to go. And then either there's a word to actually work in the lab or else to say, Hey, I've got this prompt that kind of works and saves me five hours a day. Could you make it good and get it out to everybody so it's not just a one component piece. You need the other pieces.

Jon Krohn: 02:31   Incentivizing workers rather than punishing them sounds like the best way forward for an AI positive future. My next clip is from episode number 955. This episode rounded up 2025 and marked the beginning of 2026 together with the brilliant Sadie St. Lawrence, we shared thoughts on the highs and lows of the year. That was, and Sadie who serves as founder and CEO of the Human Machine Collaboration Institute, made one disappointment of hers. Very, very clear. Right? Disappointment of the year. It's interesting because you, earlier in this episode, you used a company name and the word disappointment in the same sentence. So is that company, this

Sadie St. Lawre...: 03:15   One may be controversial. I have may think you're going to disagree with me on this, John, but I'll explain myself. And so I'm just going to say for me, actually, agents were

disappointing. And I wrote a substack this year called Agents of Disappointment. Take her

Jon Krohn: 03:30 Off the air, take her off the air. Where's the aboard button

Sadie St. Lawre...: 03:35 Of all of a sudden I get muted. You'll know why. Right.

Jon Krohn: 03:39 Alright, and that brings us to the end of the episode. Yes,

Sadie St. Lawre...: 03:41 Exactly. No, so I wrote a post this year as my best performing stack called Agents of Disappointment, but really what I was talking about was the divide between the hype of agents. And really where I saw this come into disappointment was from the enterprise companies. We have, obviously Salesforce has been talking about its agents force for some time. You have SAP, you have all of these enterprise companies who have been talking about how to implement agents into your existing enterprise tools. It just doesn't work right? Or at least a lot of companies aren't set up for it properly or in my mind, don't know how to think about how to structure them properly and what to have agents do. And so from that standpoint, from an enterprise agent standpoint, I think the hype and the practicality of the implementation, the divide between those two was too great. And so that was my disappointment of 2025.

Jon Krohn: 04:41 I totally get it, and I don't disagree with you. It makes a lot of sense to me. There's too much talk about agent AI relative to the impact that it's making. No question. I do think that a lot of it is related to people not having their data silos set up in a way or their security set up in a way where they're comfortable with it. But yeah, a lot of tinkering with agents, not nearly as many enterprise deployments, but I do think it will come. That is not my disappointment of the year. My disappointment of the year is Apple.

Sadie St. Lawre...: 05:19   I think you were disappointed with Apple last year. We back, let's look. Oh, I think, yeah, because Apple, did they announce Apple Intelligence last year or was that a this year thing? I don't know. Time is weird in the AI world.

Jon Krohn: 05:30   I think it did. No, I think you're right. I think they announced Apple Intelligence in the autumn, northern hemisphere, autumn of last year. And it was disappointing. But I mean, I guess it's kind of just like Google. It's like,

Sadie St. Lawre...: 05:43   And it's still disappointing.

Jon Krohn: 05:44   It's still because it's like a year later and what can I do additionally with AI in my phone that I actually use? Not much. I once made an and then didn't do it again, just used it built in Gen AI to send an emoji over iMessage. But I don't know. I don't really need to do that. Yeah, it seems like things like just having Siri be able to understand what I'm saying to it in the same way that opening I whisper can, I mean,

Sadie St. Lawre...: 06:25   Who's actually even done better than that is? So I will use grok in my Tesla. And so when I'm coming home from work and want to brainstorm something or learn a new subject, I just push a button and it works seamlessly and the chat mode goes back and forth and I'm like, okay, if I can do that in my car, why can't I do it on my phone that seamlessly? So I hope that next year, our comeback of the year is Apple because they've been in the disappointment quarter for far too long,

Jon Krohn: 06:56   And they have a lot of potential because of how ubiquitous their devices are. There's a big opportunity for Apple if they can get it. Right now we're jumping straight into some of the most exciting frontier tech for 2026. In episode number 961, Cisco's, Dr. Vijoy Pandey talks about distributed artificial super intelligence and how

that idea is deeply rooted in the development of human language. Awesome. So now that we have this definition of artificial super intelligence under our belts, Vijo, tell us about this idea of distributed artificial super intelligence or as I've seen you abbreviating it, DASI.

Vijoy Pandey: 07:35 Right. So if you think about human intelligence, because one thing that we have in common moving forward is the comparison bar for all of us is human intelligence. And whether you think about artificial intelligence, A-G-I-A-S-I, powerful ai, whichever definition you might have, we just talked about two definitions. One, which is economic in nature, one is technical in nature. There seems to be always this comparison metric, which is let's compare it against humans. That's the best thing that we know when it comes to intelligence. So if you think about humans and how we evolved and how in our intelligence evolved, we actually evolved intelligence across two axes. So the first 300,000, 400,000 years, human intelligence was actually scaling up vertically. So we were getting smarter and smarter. We were inventing tools, we were inventing processes, but it was limited because we weren't communicating that intelligence. So the communication was a big missing piece.

08:39 And so what ended up happening was whatever we invented and whatever processes that we came up with and the dangers that we were aware of and how we reacted to those dangers and the way we stitched our clothes together or whatever we wore, I mean I'm not an expert there, but whatever did was limited to the lifetime of either that individual or that process. And so we became more and more intelligent, but it was very, very limited. And so we were scaling intelligence vertically. But as we all know, every system including intelligence can be scaled on two axis vertically as well as horizontally. And so there was this big evolutionary jump around 70,000 years ago, it's called the cognitive evolution in humans,

where we discovered language, not just sounds, just paintings and patterns, but language. How do you convey meaning semantics between people, and then how do you convey that across humans, but across tribes and across generations?

09:48 So what happened when that cognitive evolution happened was we invented three things. The first thing being shared intent. So as a human society, we started sharing a common intent. Let's go and build this not just in the lifetime of me as a person, but in the lifetime of this tribe or this group of people. So shared intent and coordination as a result. The second thing that we invented was shared knowledge. So this is what's colloquially known as standing on the shoulders of giants. So I build a knowledge base, then you add to it, then you add to it, or you modify it and you keep doing that. And that is cumulative human knowledge. So we invented shared knowledge. And then the third thing, because of the first two is now we could do shared innovation. So innovation itself wasn't a singular pursuit or an individual pursuit, but it was a shared pursuit.

10:49 So that's what happened when language got invented and semantics got invented and you started scaling horizontally because now you are inventing as a collective instead of as an individual. And so what we are seeing, and the big thesis here is that so far in intelligence, in artificial intelligence, in artificial super intelligence, we've been building bigger and bigger individual geniuses and the framework and the infrastructure to do collective intelligence, to do distributed intelligence has been missing, and that's what we want to go after. So disparate intelligence to us is to enable to build a framework that allows for shared intent, shared knowledge, and shared innovation to happen in this multi-agent human society,

Jon Krohn:	11:41	A multi-agent human society. Sounds great. So let's hear from the people who are helping us to reach that goal. Evaluative frameworks are one way to get clear about how effective our AI systems are, and O'Reilly educator, time bestselling author and AI entrepreneur Sinan Ozdemir explains that finding a common language for those frameworks will be crucial to their success. Here's a clip from episode number 959, speaking of experimentation and research, let's jump to that chapter three that you mentioned earlier. The fun one, I mean there's lots of fun chapters, but chapter three is in particularly the good one, and in it you present your comprehensive framework for AI evaluation. So you emphasize that accuracy alone isn't enough, and so you introduce multiple metrics across different task types, retrieval, classification generation, and you stress the importance of reproducible experiments. You conclude the chapter by noting from now on we will be incorporating evaluation language into every case study. So then throughout the rest of the book, you have these fantastic detailed case studies that build on each other, and you use this common evaluation language that you introduce in chapter three throughout, which is brilliant. So when you organize this evaluation by task buckets generation, multiple choice, embedding classification, why is that separation so critical?
Sinan Ozdemir:	13:04	Yeah, so the split is usually of the types of LLM tasks, there's generative and understanding, and then under generative, there's a multiple choice and free text, meaning it's basically like auto encoding versus auto aggressive is how I try to think about that. Analogously meaning if you're talking to a chat bot or an agent, which is it's a chat bot with tools, you're asking it either to produce a paragraph, a sentence, several paragraphs, whatever, free text or you're asking it to pick from a set of options, should I proceed yes or no? Is this good enough to post on LinkedIn, yes or no? That's multiple choice. I'm basically collapsing the entirety of this deep learning

architecture into a binary classification task. So versus understanding tasks which are embeddings and classifications, which are similar to multiple choice, but just with a different architecture, each one of those has their own suite of metrics.

14:05 Because how I evaluate a child's essay on a caster in the rye is going to be different than how I evaluate the embeddings that this embedding model is producing. They're just not the same task. They're not built for the same thing. They're all in LLMs open. AI and beddings are produced by LLMs. Classification models are run for the most part by LLMs. So the evaluation is less on the model, it's more on the task that you're trying to perform and whether you're performing classification through any kind of architecture. My book from 10 years ago, principles of Data Science talked about accuracy, precision, recall sensitivity or specificity. I also talk about that at my book from two months Identical ai. It's the same classification that I'm asking an agent to do. It's the same task, it's just a different model is now doing it. So evaluation is tricky. It's the longest video I ever wrote or made was nine, 10 hours on the O'Reilly platform was evaluations because there is no one size fits all. It's what are you doing? I'm now going to walk through 20 case studies that are all very different from each other, all with different metrics.

Jon Krohn: 15:18 And so to dig into this a little bit more, you mentioned there are all these different kinds of metrics for evaluating performance. So I already said in a question a few minutes ago how accuracy isn't enough. In your book, you emphasize how using precision recall and where you're trying to rank results something like mean, reciprocal rank MRR, using those metrics together because each exposes different failure modes of a model. Do you want to tell us a bit more about that?

Sinan Ozdemir: 15:49   Yeah. So precision recall is probably the more I would say usable metrics for most people. Meaning I'll say it this way, if you ask an LOM, you give it a LinkedIn post and you say, is this going to get a lot of engagement on LinkedIn? And it says, yes, okay, great. You post it, it doesn't get a lot of engagement. That model had a false positive. It told you yes, but really it was no. When you care about false positives a lot, when they are expensive to you, you care about precision. Precision is the measurement of all the times the model said yes. How often can you trust it? So when the model says yes, go ahead. How often is it correct in saying yes, that's precision. So when you care about false positives, precision is your metric recall is kind of the opposite. Recall is of all the times it should have said yes.

16:52   How many times did it, so a false negatives are expensive to you recall is the metric you care about because if the thing says this is a terrible LinkedIn post, but you post it anyways and it gets a lot of engagement, that's a false negative. It didn't want you to post that and recall is a measurement among other things. A recall is effectively a measurement of how many false negatives that you're seeing out of this system. And that was a pretty dense explanation for two. Honestly, one of the simplest metrics in machine learning, and it kind of goes to show that the conversation around evaluation is not always as simple as here's the fraction that you care about. No. Before we get to math, what do you the human care about, what's expensive to you? If you say this factory part off the line is good, but it's not good, is a plane going down or is someone's light going to break? How expensive is a false positive to you? If it's expensive, precision is the thing you need to look at. Recall shouldn't matter as much. I'll happily throw a part away on accident. At least if I know everything off the line is going to be right. Precision matters the most. So again, it always comes down to not just the task, but even the risks of failing that task

Jon Krohn:	18:11	From risks. We return full circle to incentives with my final clip, which I'm taking from episode number 957 in this episode, who is the co-founder and CTO of Excel data, a Bay area AI startup that's raised over a hundred million dollars in venture capital. Talk to me about how to find and keep the best developers and data scientists in their jobs. The kinds of things I was going to ask about in my next question, which were, because you've said in past interviews that a talented programmer is looking for meaning in their day-to-day that highly paid engineers still just complain about their jobs. It's funny to me to think that somebody who gets a hundred million dollars signed bonus said meta is then just like, oh man. But I'm sure that happens. I don't know where the stat is at today, but when I was doing my PhD, something like 15 years ago, I attended a lecture on the economics of happiness.
	19:07	Just for fun. I just went to this lecture and at that time it was showing that in the US, if a household was making over something like 80,000 or a hundred thousand US dollars a year, that is the happiest you can be making more money beyond that point, did it make people happier? Now with inflation, the numbers are probably a bit little bit higher, but directionally, I think this kind of gives the idea that you're explaining, which is that beyond having your basic needs taken care of and knowing that you have security for you and your loved ones, the extra money beyond that could end up being a hassle.
Ashwin Rajeeva:	19:43	Yeah, it is. It is. And it's also interesting, right? I mean there are studies on developer productivity, right? And you would see that the numbers are insane. I mean, people talk about how an engineer, a software productive maybe four hours a day or three hours a day and the rest of the time is spent in meeting planning, whatever. It's right now I feel that even if you take two hours for meeting, you still are leaving a lot of this time out. And if

you think about it, what better privilege can someone have than to sit in usually a great office on a laptop without having to move moving is optional and then get paid top dollar for it. And most engineers then leave their jobs. It's not just people leave accelerator, but people leave all sorts of companies and there has to be a reason for it is that most people would be happy because knowledge work is something which has to do with creativity.

20:45   It's hard to sit in one place, realize that the work which was presented to you or asked of you could be done maybe in two hours. And then you've got to sit and find something to do and it's good for a few days and you spend some time, but after a while you start getting this feeling that, hey, what am I doing? I have supposed to do something better. Let me find a mission which resonates with me and my work and my philosophy of it. And so I think making sure that no matter what the business environment you are in as an executive, the engineers or the r and d teams believe that fundamentally we are in the business of innovation in this field. And there are very few fields, whether it's something data management, even something as boring as the enterprise content management. I'm sure that there is innovation that could be done, new ways of doing things and people should believe that they have the freedom to do it and they're not just dictated by quarterly plans.

21:48   And this, I think if we can provide an environment like that, then new ideas come in. And for us it's worked out because for a company of our age and size, we have a lot of, let's say, capability that we have built over the years, whether it's to do with ODP, aoc, we have a pulse monitoring system, we have a DM, we are working on the next version of our platform, which will be released in May. And so that is what allows us to do it is where people believe that hey, in this field of what the company

has chosen data management, there is innovation that can be driven through pure engineering work. And that's what drives people.

Jon Krohn: 22:31   Nice. I like that. How do you say in an interview, do you think you have a way of telling whether somebody's going to be passionate about a technical infrastructure heavy mission like data management at Excel data versus somebody who's just coming to collect a paycheck?

Ashwin Rajeeva: 22:47   I think it is easy to tell in some sense. Of course we have made mistakes there as well, like everybody else. But I feel once you start talking to people, and this is what I felt, I mean I've always felt that management in the technical field can only be done by people who have been in the trenches to some sense. And I'm sure there are models everywhere else and which are different. And people have seen managers who work extremely well without actually being on the field. And so the number one thing, at least when I look for potential hires, is to see if they can build things. And it doesn't have to be working on some data problem. The question is, can you build if you are given a problem, and it could be any problem, it could be something like, Hey, how do you design, let's say a e-commerce warehouse, how do you design a logistics system or any other business problem? And then can you translate it to something which you have learned?

                  24:12   You probably know, go Python or Java. Can you put something together which represents a real world problem? I think if you can, then those are the people who bring the most value, who can actually look at a business problem and then convert it down to what they know. And that's what technology are good at. Of course, this is for slightly senior people. I think for people who are just coming out of college, it's purely based on potential saying, Hey, some of it is your scores and your background and some of it, Hey, how interested are you

into doing this? And then you take a bet and maybe after a few months you decide, but for most senior people, I would recommend checking if they can build things.

Jon Krohn: 24:53   Alright, that's it for today's ICYMI episode, to be sure not to miss any of our exciting upcoming episode. 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.