



SDS PODCAST

EPISODE 794:

EXCITING

(AND FRIGHTENING!)

TRENDS IN

OPEN-SOURCE AI



- Jon Krohn: 00:05 This is Five-Minute Friday, on exciting trends in open-source data science.
- 00:19 Welcome back to the Super Data Science Podcast. Today's episode was filmed live on stage in May, at the New York R Conference. I hosted a panel there that consisted of four data science icons. We had Drew Conway, who is head of data science for private investments at Two Sigma, one of the world's largest hedge funds. We also had Jared Lander, who's adjunct faculty on Statistical Programming at Columbia University, and author of the bestselling book, R For Everyone. We had Emily Zabor, who's a biostatistician at the Cleveland Clinic's Department of Quantitative Health Sciences, and JD Long, who is VP of Portfolio Solutions at Renaissance Reinsurance.
- 00:55 Today's episode will primarily be of interest to anyone who's keen to hear deep experts opine about the future of machine learning and AI. In today's episode, the panelists answer a question from me about what they're most excited about, or most concerned about, with respect to the fast-moving open-source AI developments that are transforming our world today. Ready? Let's jump right into the panel.
- 01:18 We are at a time in history where we're seeing really powerful AI technologies. We are seeing AI talked about by the public for the first time in the last couple of years. In a lot of ways the open source communities, and obviously the libraries, that have been developed in R, in Python in particular, these have allowed this AI renaissance to occur. I am a techno-optimist. I believe that we are on the precipice of much, much bigger things, and that those great things are accelerating. I think that we have a lot of wind at our backs, in the sense of data costs, data storage costs plummeting all the time, data



collection increasing exponentially year over year all the time.

02:10 Compute costs going down a lot, exponentially all the time. With these kinds of winds at our back, we have more and more intelligent data applications at our fingertips, in our browsers, on our phones, in our cars, in devices increasingly all over the home and the workplace. This is a big question, but what are you excited about? What are you concerned about, in terms of data and AI applications in our lifetimes, or maybe even in your children's lifetimes? Drew?

Drew Conway: 02:47 Sure. Well, I guess in the short term, for excitement, and actually I think for folks who are here who are listening to Hillary's talk, she really did a great job in classifying the challenges of general purpose use of transformer and GPT technology, this idea of accessing just the median output of stuff. In my professional life, we use a lot of these tools really as a way of creating more efficient workflows for more traditional modeling. Like finding these well specified problems that have a testable right answer, that we have confidence in the output, and then we can take those answers and say, "Well, it would be much more efficient for this machine to do it." Even with some stochastic output being wrong, we can still use that technology. That allows us to have higher order thought on what we might want to do with that, say, features that we generate as a result of that. So, that makes me really excited because I think, like any other tool, it just allows you to expand the canvas of stuff that you can do with them.

03:55 I mean, certainly on the negative side, I think we just don't really... The technology is like this, particularly I think on the generative side, that have this pernicious influence of what people believe or don't believe, or even again in a professional application, it's so much easier

now to apply these things without having any sense of how they could blow up in your face. Maybe not even caring until it's too late, and then you're really in a bad situation. So, I think that the pendulum will swing wildly back and forth for a while. I guess the thing that I'm most uncertain about is what will become the new normal around this stuff. Without getting into too large of a conversation, what role governments will play in that, and what role private industry will play in that, and academia. We don't tend to be very good at that kind of trilateral negotiation.

Jon Krohn: 04:53 Great answer. Jared?

Jared Lander: 04:55 Also for the excited part, embeddings, because it's like it's trying to make a mathematical understanding of the world. I've seen some use case recently, if you're trying to describe a bunch of symptoms, you might say you stood up and you felt faint. You might say the room was spinning, you might say you were dizzy. It was all vertigo. Now if you can map those all together and have embeddings, they should all have a similar distance to each other, then be able to say, "Okay." Well, this then gets generative and talks about vertigo. It's a way of helping create understanding. Whether that's translating languages or just translating concepts, I think that is really cool, if an embedding can become a true deep understanding of the world. I think that is really, really awesome.

05:39 The thing I'm worried about is, and I'm totally taking this from someone else. They said that all this AI is automating away all of our creative work, leaving us time to do manual drudgery. Where in reality, we want to do the other things, but get away the manual stuff and let us be more creative. So, I don't want that to happen. I want to open up the creativity.



- Jon Krohn: 05:59 Great answer. Yeah. Emily?
- Emily Zabor: 06:02 Yeah. It's a good question. I tend to be an elite adopter of new things. It's both a personality trait of mine. I'm really risk averse and also in the field I work in, in medicine and biostatistics, we're a little bit, I think, less tech-savvy and a little more behind the times. Partly because of data privacy issues and some other things. I'm curious to see. For sure AI is coming to medicine, and I think it's going to be very powerful tool actually in medical applications and actual patient settings. Also in dealing with electronic health record data, which is a completely missed source of information right now in many, many ways because doctors write their notes into these charts.
- 06:43 It's not in a variable field where we can get it into a data set that I can analyze, but the information's in there. So, large language models are a perfect fit for something like that. I think there's going to be some great applications in medicine that I'm excited about. I don't know what they're going to be, and I'm personally just waiting to see. I'm not doing anything right now, but I'm excited about that. On the other hand, I'm like, should I even be saving for college for my kids? Is there going to be a job market in 15 years when they're going into the world? I don't know.
- Jon Krohn: 07:16 Lots of exciting things happening in medicine for sure. Yeah, could potentially have a big impact on our lifespan, quality of life as we get into old age and dealing with the big retiring population coming up. Hopefully AI can help us a bit in medicine. JD?
- JD Long: 07:35 We all think in analogies, or at least most of us do quite a bit. I find I do that more as I've gotten older because I've got more things to make analogies with. I'm reminded of my optimism about the internet in the early 2000s, but it was also naive. My optimism, I love old machines and I had an old motorcycle, I've had old Porsches, and the

online communities in the early 2000s were just amazing, because I could get the old manuals and there was all these other gear heads who were working through the same problem. I could post a picture of something and somebody would tell me what the problem was and how to learn how it worked and fix it. I was just like, I felt like I was living in some sort of renaissance of information, because just 20 years earlier, 10 years earlier, if you had an old machine, and God help you, if you had an old British or German machine, and you lived in a rural area in the United States, it was very hard to work on it, get parts, figure out what was wrong.

08:38 The internet totally changed that. I felt like, wow. For people like me who are makers and tinkers and everything, this is an amazing world of renaissance of everything. I had no appreciation of how also it would allow crazy people with really messed up ideas to all find each, and through the process of selection bias reinforce all the worst qualities of humanity within each other. I never saw that coming, never in a million years would have spit-balled that that's what was going to come out. It did. But we still have the other, so it's not like it turned into the bad thing, it's like it's the good thing and the ugly thing. I think we're going to see that with the AI as well. There's going to be a whole bunch of stuff and where some of us are just going to be like, "It's amazing. It takes a bunch of drudgery and a bunch of stuff."

09:26 Then there's probably also going to be some awful stuff that I'll sit up here and look back and go, "Yeah. I remember that conference when I said there's going to be some [beep]. Well, there's some [beep]." But I don't know what it is. That's what my frustration is, I don't have enough foresight to see through the veil of uncertainty of exactly what it is. But I guarantee there's going to be some messy stuff as we progress. I just hope that we're savvy enough and collective enough to navigate it,



because we've had a mixed outcome on the internet. We could argue, net good or net bad. I think net good. But it's bumpy. We've got bumps ahead, y'all.

- Jon Krohn: 10:05 A year ago we had drones being used for the first time in battle without any human in the loop in Ukraine-Russia conflict, in order to help those drones be able to get through jamming, that the Russians were doing. So, that sort of thing could potentially, you could imagine a few bad outcomes from weapons being fully autonomous.
- JD Long: 10:29 You can go read Daniel Suarez's books, if that's appealing to any of you. He had a whole series of books around some of these things.
- Jon Krohn: 10:36 There you go. All right. That's it for today's exciting and slightly frightening episode. If you enjoyed it, consider supporting the show by sharing, reviewing or subscribing. But most importantly, just keep on listening. Until next time, keep on rocking it out there. I'm looking forward to enjoying another round of the Super Data Science Podcast with you very soon.