SDS PODCAST EPISODE 949: WHY AI KEEPS FAILING SOCIETY, WITH STANFORD PROFESSOR ALEX "SANDY" PENTLAND

Jon Krohn:	00:00	The Soviet Union collapsed because of AI. And today's guest, professor Alex Pentland, argues society is on the same crash course today. What the heck can we do to avoid AI Armageddon? Welcome to episode number 949 of the SuperDataScience Podcast. I'm your host, Jon Krohn. I'm joined today by a blockbuster of a guest, Alex Sandy Pentland, a distinguished professor at both Stanford and MIT, who pioneered the field of computational social science. In today's exceptional episode we focus on is brand new book Shared Wisdom, particularly the biggest risks to society from AI and how to mitigate them. This is a special episode. Enjoy this episode of Super Data Science is made possible by area Dell, Intel, Fabi, and Anthropic. Alex Sandy Pentland. Welcome to the SuperDataScience Podcast. It's great to have you on the show. How are you doing today?
Alex Pentland:	00:55	Pretty good. Very good. Glad to be here.
Jon Krohn:	00:57	Excellent. And yeah, so your full name to get at that right off the bat is Alex Sandy Pentland. People can find you very easily. You've got great SEO with all three of those names together. Okay.
Alex Pentland:	01:13	Sandy is a nickname, the Sandy is, my dad was also Alex, so I had to get the nickname. And there you are, Sandy.
Jon Krohn:	01:21	Nice. I like it. And so I think that is generally what I'll be calling you in today's episode. It feels very familiar, so I appreciate you welcom me into that. So you are an exceptional guest for us to have on the show. I've seen you keynote at the Open Data Science Conference before. You've been a professor at MIT for nearly 40 years, pioneering research there at the intersection of social sciences and computer science. And more recently for over a year now you've been a research fellow at the Stanford Institute for Human-Centered ai, HAI, which is a leading institute as well, just like MIT, for this kind of

research, countless papers, something like you have dozens of former PhDs and postdocs that now have prominent positions at universities. You've spun out over 30 businesses from entrepreneurship programs that you've been involved with, really an amazing guest to have on the show. And you have a brand new book that we're going to be talking about as well. So your book is called Shared Wisdom. Tell us a bit about the book and why you decided to write it.

Alex Pentland: 02:29 Well, as you mentioned, for a long time I've been looking at how humans interact and make decisions and using sort of a big data approach to it, looking at tone of voice first we had little wearable things that we did. We did all the wearables stuff back in the day. And then looking at how people move around and mix with each other and so forth. Just sort of say, well, what makes a good neighborhood? What makes a good conversation? What makes something that's really innovative? And I think we've been able to bring it all together and it's really interesting because it has a lot to do with what sort of AI is going to be successful and what is not, and what is it going to do to corporations and countries and democracy. And I think I have a really sort of good understanding and a unique understanding because I'm looking at it through the lens of how people work. Because after all, at the end of the day, most of this is going to be used by people. It's going to compliment them or attack them, the two. And you want to have design of the bigger system, which is AI plus people, plus all the other people out in society. If you have an understanding of how all that works, then you can build things that really rock. If you just try and build it, that's like as if people don't exist or people are just logical engines, you're going to make mistakes that are bad mistakes.

Jon Krohn: 03:58 Right. Yeah. We've got a quote from you where you argue that AI systems, when they fail, they rarely fail because of

inadequate algorithms, but rather because of inadequate models of society. Absolutely. So yeah, what does that mean and what does it Yeah, yeah.

Alex Pentland: 04:16   Well, so lemme give you an example that probably most people aren't familiar with, but it's really prototypical. So back in the late fifties, early sixties, that's when AI started, John McCarthy, Marvin Minsky got together and said, what name could we pick? Which was the scariest so that the government would give it us lots of money. And they came up with AI. And incidentally, I have Marvin's endowed chair when he retired because we sort of compatriots there. The thing that really sort of took off back then is a thing called optimal resource allocation. It's basically linear constraints and linear models. So it's in things like Excel, it's everywhere. Anything that's doing path planning, anything that's doing any of those sort of optimizations actually makes use of the math that was invented there. And famously, the Soviet Union tried to run its entire economy on that form of ai and it actually works.

    05:25   This is not a problem. You can do it in all sorts of things, but you need to have good data. Does that sound familiar? And you need to have a situation where your equations and your models are up to date. And so what they discovered is that first of all, people would lie about the numbers. And then second of all, things change over time and there's these unexpected rare events that come along, something crashes or whatever. And those would always just bring down their model completely. And it's that using ai, that sort of AI to run the Soviet Union was a major contributor in it falling apart. Okay, well so now go ahead. Another generation and you've got expert systems. People say, well, the math is not that good, but we have these smart guys who know how to run a business. Let's just take all their rules.

06:18   And again, both of these things are actually everywhere now. So it is thought that that optimal allocation, the linear systems and constraints is the most common ation in the world. Even today, expert systems like rule-based systems, they're everywhere. Come on, you don't even think about it. But they do run all the sort of big corporate systems. What do you have to do to get us something approved? What do you do to audit things? The tax system is an expert system and it does a lot of good things. Okay, it's cheaper to run, but it's not very flexible. And if you look at different communities, different groups of people, it's built for the average, it optimized for the average, which means if you're a little bit off the average, it's not that good for you. And if you're a lot off the average, you're really not good for it.

07:15   And so it's everywhere and it's sort of why we don't have local clinics anymore. It's why we don't have local banks anymore. It's why we do this bowling alone stuff people may have read about. It's sort of destroyed those community structure because everything got centralized and not really all flexibly centralized. And then everybody here in this podcast is familiar with machine learning. Machine learning is sort of the same except that again, you have to have the right data. It's backward looking, not forward looking. And so you get echo chambers and social media, you get people doing all sorts of stuff that is two heartbeats behind and kill your company. So you need to design these things in a way that fits with the way the people who are using it and the people who are buying things and the bad guys all are taken into account. That's the point of that comment. Sorry to go on for a long, but

Jon Krohn: 08:25   Yeah, no, no, it's your episode. You can go on as long as you want it. I don't even think you went on

Alex Pentland:	08:32	To, it's like we keep doing the same stupid things over and over again and maybe we ought to think about that this time.
Jon Krohn:	08:39	So it sounds like you might be making a parallel there between the way that the Soviet Union fell apart as a result of the reliance on the kinds of AI systems that they had then. And maybe are you suggesting that this system that we have today could fall apart because of the AI systems that we have today?
Alex Pentland:	08:54	Oh yeah. And it'll be the same sort of issues, bad data. The fact that it doesn't take into account the fact that things change quickly, the fact that it doesn't really pay attention to human nature and how people work. People are not logical, come on. But we do know a lot about how people make decisions and if you build something that is complimentary to that, you get better system performance system, including the people, right?
Jon Krohn:	09:27	Right. Yeah. So I have a quote from you that's humans aren't rational. Instead they're what you call social forger who learn by watching others. So they're kind of picking up social behaviors from cues, seeing other people. I see that in the way that my puppy learns as well.
Alex Pentland:	09:45	It's true of almost all animals and certainly all social animals. I mean just to illustrate it, nobody ever reads manuals, nobody ever reads the whole newsletter when they get a newsletter. What we do is we watch what other people do. We ask them one or two questions that we don't understand and that's how we sort of pick things up. Danny Kaman, who's a Nobel Prize economist who invented this fast and slow thinking stuff people have heard of estimated that probably 95% of all the things we do come from other people from the culture around us. This is the way we do it here. We talk about company culture, we talk about the culture of the country. And

what that means is that there's certain ways you do things and certain ways you think about things. And we've all sort of agreed that this is the right way to do it.

10:38   That's our wisdom. Doesn't mean it's right. Don't be suckered by that. And most of what we do is that, and then we add a little bit of our own circumstances to it for that other 5%. Other 5% is really important. What makes you, but it's not the majority. And so what you need to do when you design systems is you need to think about the context, the human context of somebody who's making the decision. And this is where things like the way social media is designed, thinks of people as just individual, and then it says, oh, well we'll just use machine learning to get people that have similar feature vectors. They don't think about the community aspect of it. So you get echo chambers. Same thing as we've seen that in social media. People decided you could make a lot more money if you allowed followers and a little bit of advertising.

11:41   And so that's what we got. And the consequence is that the only people you really hear on social media are these sort of very loud voices that have done everything they can to be heard. They're trying to be violent and crazy and this and that. And we've been able to show very, very definitely. That's where polarization comes from. You get rid of these big voices and you discover that people actually pretty much agree with each other. And so if you're building a decision system within a company or wherever, you can't have these super loud voices, otherwise you're just going to get craziness going on. Most people are pretty sensible. If you sort of focus 'em on that and prevent the big voices we find you get much better decisions. Another thing that sort of comes from this is people think, oh, well, human decision making, that's only approximate.

12:44	And really what we want is a logical framework and all that sort of thing. Well, in the experiments that we do, and we do this with people who trade $10 million a day, so these are people who are super quant, super expert. This is what they're paid to do. If they pay attention to the other traders, they do better, reliably better. So all those equations mean trading currencies for instance is a real traditional, you could write down the equations, but you have to pay attention to other people to avoid edge risks. Unusual things that are changing. And actually we probably all remember 2008 when the financial system almost fell apart. The core thing was they had an equation and they didn't notice that that equation had a tail risk. So unusual events, and when an unusual event happened, it almost took down the whole system. The people who listened to other people who were also looking at actual human behavior in their equations came through it. The people who are just following the logic got creeped.
14:08	We are not bad at making decisions, particularly in small groups. It's good for us to listen to the numbers, absolutely. But if you just follow the numbers, you lose the context, you lose the changeability of the situation, you lose the possibility of rare events and you tend to come up with systems that don't perform as well as if you do take the sort of human context and thinking into account. That's what the book is about incidentally, right? It's like how do we make systems, AI systems that are really robust to weird stuff where they fall apart, you get something that's weird or an edge case, things could blow up in very unexpected ways. We hopefully all know that. And that's why some of this human reluctance to do new things, it's attraction to where unusual stories, some of these traits we have are actually extremely valuable in making sure that systems don't blow up.

Jon Krohn:	15:17	What is some of your practical advice in your book, shared wisdom on how we can be avoiding some of these issues? I mean some of them sound very difficult to deal with. For example, the polarization that we have in society, these big voices that are out there, you talk about how we shouldn't have them, but they'd kick up the biggest storm of all if you just got rid of them
Alex Pentland:	15:42	And they're there because of the economic model underlying these things. I want to talk about two different approaches to doing this. Well first is an individual approach and the second is a group approach and the polarization comes into group approach. So we have a project that's called loyal agents.org. It's all one word, no underscores or anything. And it's an attempt to make AI agents that work for you personally. So we're rushing towards this world where everything is AI agents, but the questions come up, who does that agent really serve? Is it her serve Amazon or does it serve you right? Also, what about edge conditions? If you put in something unusual, does it occasionally blow up and give you weird stuff? Well, yeah, actually it does. And sometimes they'll even do things that are illegal. You have to sort of have real guardrails. They're not good at guardrails at all.
	16:50	And then when you get networks of agents all talking to each other, you find they can get lost pretty quickly. Lost means they sort of forget what they're supposed to do. They begin to not follow what you'd like to see them do. So how do you make sure that these agents work to fulfill the intent that you have and do it in a way that's legal not going to get you in trouble and reliable. Those are the questions that current AI just doesn't really answer. And so that's what we're doing is with the Stanford Law School and so forth should take a look at it. It's really interesting because make you think about how did you design AI agents and networks of AI agents that aren't going to get you fired, that sort of thing, or your company

go bankrupt. So that's one area, and let me just illustrate it by a really simple example.

17:49   I was on a panel with the CTOs of some major companies and we were talking about how hard it is to get quality data and how changeable things are and how you have to keep really up to date and keep track of what's happening. And somebody mentioned that they had built AI buddies for all of their employees. So what is an AI buddy? AI buddy is an on-prem ai. It's not a very smart one, it's not an edge one at all. But what it has done is it's read all the manuals that you haven't read and the newsletter that you never really get to, and it also has news from what other people are doing around the company. And so what it is is it's keeping you in the loop. We have all these things, but people never use 'em. Just we're too busy. You were talking about some of this earlier.

18:40   When people watch this podcast, it's when they have a little more time. So time is an important thing. So you can make a little AI that pays attention to what everybody else does and helps you says, oh, I see you're doing this over in this other department. They do it differently. You might want to think about it not telling you what to do. If you're in a situation where the AI is telling you what to do, you have made a mistake, you need advisors, you need something to help direct your attention, your context and so forth. But don't let it make decisions, at least not in the next several years. Not that reliable. They're not bad, but they're not that reliable. And so this notion of AI buddies is trivial to spin up. I mean this is just a few days of work by a good team and you have all the data resources already, but it really changes the culture of a company because now people know what everybody else is doing.

19:40   It's not like you're working in this isolated place anymore and it's particularly good for work from home and it's particularly good for distance places. So you have an office in Timbuktu or wherever, how do they stay in the loop? Well, this is brilliant for that and it's very, very cheap and very, very easy. So this is something that really helps do things by aiding human memory and human attention. And I said what people do when they make decisions is they look what other people do. Well, it's helping you do that. It's not telling you what to do. It's helping you be a better human at finding out what are the options. So that's one thing. The second thing I want to talk to you about or just to sort of answer your big question is a thing called deliberation.io io. So these are all open source, all the sciences there.

20:41   You can get as geeky as you want, but what it does is it is sort of like Twitter slash x except that there are no big voices. You just contribute things. And then what it does is it does a visualization of what everybody thinks. And you know what that does. That just destroys polarization. Polarization comes from having people, you start to think about what does the other side think? But when you ask that question, the thing that immediately comes to mind is the crazy guy with 5 million followers. And indeed there are crazy guys with 5 million, but most people agree on even very contentious things like if you look at things like abortion or gun control, most people in the United States actually agree and we're reasonably liberal. But then there are voices on the end that will, if you allow them into the discussion, that's the end of discussion.

21:45   Now it's a polarized fight to the end. So you set it up like that. And then the other thing we have in there is AI as mediator. So a general principle of what we try to do is not have AI's tell you what to do. They want to remind you what other people are doing, what could happen, things like that. So they're like a good librarian more. But

in this case it's a little bit more than that. It's something called Socratic dialogue. So the AI, in addition to visualization of what everybody is saying will have little summaries. I hear people saying this, but then some people say that that's not adding its own content. There's a little bit of room for bias in there, but it turns out they're really good at that sort of stuff. And so what it can do is listen to a large number of people and say, here's the themes that come out.

22:42 It turns out that if you reflect that back to people, they do just dandy in experiments. We have people like the AI summaries twice as well as they like human summaries. Think about that. And this is their normal people. This is not specialized experts and it's not a fancy AI at all. You can do this with a really cheap ai. It's just trying to keep people on track. That's really all it's doing and trying to make them aware of the context. We did this with Washington dc, the city, not the Senate and people like that. And we asked them, what do you want for AI and city government? Sort of interesting. And the thing that none of us experts expected is what the people wanted. And we're talking about people who are two jobs, three kids, they haven't got a lot of time, they're real people.

23:41 What they wanted is they wanted AI to defend them against city. They wanted a personal AI that would help them deal with all the regulations and find opportunities and fill out those damn forms and et cetera, et cetera, so that they could get onto the important things in life and wouldn't miss out on opportunities. So when we did this big group discussion with the people of Washington dc, what they wanted was personal agents to help them get by. So I think those are two themes that are real good examples of what you can do. So the deliberation thing, I mean, just think about it. Imagine that all your meetings were twice as good as they are now. Maybe they only go on half the time. Maybe their decision is twice as good,

but you like them a lot better than you do now. It's the number one thing people complain about in companies and in other organizations, these the endless meetings. Well, now you have a very simple way that doesn't intrude on the human agency and our sort of humanness to make decisions faster, better, more inclusively. Well, okay, this is just something that we put together. It's all there. It's on the web, et cetera. It's not GitHub actually, but same thing, but just think what we could do by taking this sort of approach of helping people think better rather than trying to replace people. So I went on a bit long there, I apologize.

Jon Krohn: 25:20   Well, yeah, to recap quickly, the two main ideas are having AI buddies and this kind of deliberation.io framework where these big extreme voices aren't amplified as much as they are in our typical social media system. Those are two of the ideas there. Something related. So you provided some society-wide solutions there. Most of what you were talking about was kind of very broad. An interesting specific point that we've noticed you talk about before is in academia where experts often cluster in narrow citation networks and where rewards the reward structures that we have trap researchers in local maxima. And so would you say that these same kinds of solutions like an AI buddy that can be giving you tips on what people are doing in a different domain that could be relevant to what you're doing?

Alex Pentland: 26:24   Yeah, so we've built something and you have to look at sort of community deliberation with my name, but there's an archive paper about exactly this. These are tools we're building. It's actually pretty simple. It's an AI buddy that says, where is in the space of ideas around you, A place that nobody else is looking. So there's two things you need. One is people aren't looking there. So you can see that through the citations and the data. It's not a subjective thing, it's an objective thing. And you can also

say, Hey, but this blank spot that nobody's looking at has lots of people nearby so that if I can figure out something to do there, then what it's going to do is it is going to be of interest and affect everything else. And it turns out those blank places are the places where you get superstar papers, the papers that are cited a bajillion times and change the science.

27:25   So that is an AI buddy that's helping us with our blind spots. It also turns out that you can build the same AI buddy and run it on patent databases. And what it will do is find you places that there's lots of people patenting things, but not in the center. And so you can say, oh, what am I going to do in the center here? This is likely to be commercially very, very valuable. And if you put the two together, it's like, okay, where am I going to put the research money and where am I going to put the development money? And then the third one that's really funny is that law works the same way. So when they do legal decisions, they cite other people and you can figure out areas of the law that are going to have a lot of, let's say churn. And you probably want to stay away from them, maybe you want to dive in, but I would stay away.

Jon Krohn: 28:20   Great. It's nice to have your insights on what we can do society-wise with AI buddies as well as at the academic level. It's incredible all the different initiatives that you're associated with. Another comment that we came across from you or another perspective that you have is that you've argued that data have become a primary means of production, your words, and that just as labor and capital eventually gained unions and community banks, you argue that communities will need their own data unions to counterbalance corporate concentration. I had never heard of an idea like this. Do you want to tell us more about it?

Alex Pentland: 28:57 Well, so first of all, it's pretty clear that data is a means of production. And I will cite sort of the loudest voice in the world. Xi Jinping was the head of China who says this says, so that's really weird. There's this guy who's Marxist. That means that his whole theory is around capital and labor. And he's saying, no, no, no, no, it's no longer capital and labor. It's capital labor and data, and if you control the data, then you can manipulate the other two. So that's what the head of China is saying, not just me. He apparently did get it from me, but that's okay, right? Happy to give away ideas.

29:43 And of course that's where all the hyperscalers come from too, right? It's all huge concentrations of data. And then you ask, well, so what happens to people like me or, well, we are sort of at the mercy of all these people who know a lot of stuff about us. We don't know what they know about us and they're trying to get us to buy things, right? It's one of the reasons that personal agents are interesting because personal agents don't have to share data to get results. They can look at what's out there, they can ask specific questions, but they don't have to tell you, oh, I'm a male this age, I make this much money. It can sort of be a lot more private inquiry sort of model. So that's actually really interesting that it is a chance to take back your data. The reason the loyal agents.org thing is with consumer reports is because when you get a group of people who are using their data to figure out what things are scams and what things are a good deal and what works and what doesn't work, then you have the power to be able to begin to compete with fairly big people like Walmart who know a lot about you, but now you know more about you than they do.

31:04 And by sharing among the community the cooperative, you can operationalize that to get better prices, avoid scams, all sorts of things. This is not a new idea at all. So in the mid 18 hundreds, the only sort of banks in this

country were in New York City and Boston and Philadelphia and all the people who lived to the west were getting killed. They had no financial access except on exorbitant terms. And so all these agricultural communities started agricultural banks cooperatives and broke the monopoly of the eastern banks with something that looked impossible. But by pooling their resources, they were able to do this. Those same co-ops incidentally are the ones that electrified America. You hear about government initiatives to put in electricity and stuff like that. But the majority of our electric grid came from communities building their own bloody grid because the government wasn't doing it and they didn't want some company to own their electricity.

32:15   So they built their own. And it's where we have credit unions today, that's where they've sort of evolved. After a while, the laborers noticed that this was happening and said, well, gosh, we're going to have our own unions. And these were unions of labor. And famously this industry beginning of the 19 hundreds, and eventually people changed the laws and said, no, you can't require kids to work and you can't require people to work more than a certain number of hours a day, et cetera, et cetera. And so there's this battle that happens between these means of production, and that's what Marx was about. But now we have the battle with data, and I think the same solutions that worked with capital and with labor will also work with data. And I said this sort of AI agents type of thing where you get to not share your data, but just share questions and answers. If you gather people together into a co-op, you can have something that's much more powerful than you alone can ever be.

Jon Krohn: 33:17   I like that. I like that you keep coming back to the agent theme since it's so topical. Yeah, I don't even have to what guide you there?

Alex Pentland:	33:24	The thing about the agent thing, right, is yeah, some of it's hype, maybe a lot of it's hype, but it's a different way to organize compute and data and money also for that matter. And whenever you get that change, there's an opportunity to do things in ways that we will all enjoy. It hasn't happened for two generations or a generation at least now we have an opportunity to sort of change the basic blueprint of the web. That's something to jump on.
Jon Krohn:	34:01	Yeah. And so beyond agents, you've shaped the architecture of global data governance. This is going to be a bit of a long question, so you're going to have to bear with me here, Sandy, but you've co-led the World Economic Forum discussions in Davos that produced GDPR. You served as one of the UN Secretary General's, data revolutionaries who forged transparency mechanisms and the sustainable development goals. Your work spans advising the UN G 20 OECD, founding the trust data alliance building open source infrastructure that makes AI and data safe, trusted, and secure. And now we face unprecedented threats from AI generated social engineering. So things like phishing, computer viruses, misinformation engines that we've already talked about earlier in this episode. And in the final chapter of your new book, you propose a solution which is, well, a whole bunch of solutions beyond just agents. So there's things like what you call post accountability through open audit trails, strict liability, trusted execution, and continuous monitoring rather than preapproval gatekeeping. Do you want to highlight any of these solutions? I'll tell you first
Alex Pentland:	35:13	Of all a little funny story. So I was helping this thing called the Club de Madrid, which is all the prime ministers and presidents from Europe. And we were talking about governance of ai, what are we going to do? And a lot of the people there were the heads of the EU commission or the regulators, and I was saying, the

solution I'm going to tell you, and they said, you're too Anglo-Saxon. We're going to go another direction. So what you're about to hear is, but it's not just Anglo-Saxon. I think it's something that works in lots of parts of the world, not so much Europe. The idea is that look, unless we have sort of consistent regulations about AI and its use and performance, and we're talking about things like is it self-dealing, is it biased? Is it doing the right thing at all? Unless we have consistent regulations about that and ways to check, it's just not going to happen.

36:17 And consistency is not within the country. Consistency is across the whole world. People are building data centers everywhere in the world, and your query is going to happen in wherever the energy is cheapest. That may not be in the continental us. And the things that you buy just like today don't necessarily only involve US production. There's logistics chains from overseas. And so what are the systems we have that actually work consistently across the world? One thing is purchasing systems. When you buy something, you buy it, you buy it with some amount of money, you can convert it to other money. If they screw up, you can sue 'em. Sometimes it's not so good, but at least people agree what counts as being screwed up and what doesn't count. If you don't get the thing, they owe you the money. Now you may have good luck finding that money, but at least they agree that that's bad.

37:16 Whereas today in ai, nobody agrees about anything. But what we can do is we can treat AI the way we treat money or trade and say, well look, you don't give away money without writing down who you gave the money to or where the money came from and what it was for. You have an audit trail so that in case bad things happen, we can find out what really happened. We need the same thing with ai, not we have logs of all sorts of stuff on computers. Anyhow, not a big deal. But you need to have

logs of what your AI is doing to who with what data. And if you have that, then you can find when you're running off the rails, you can detect which agents are better than other agents and you can do this internationally. And so what you could begin doing is you can say, well, look, country X, your agents really suck.

38:11   Here's the audit trail. And so they're going to raise their standards. Now you say maybe they're going to fake the audit trails. There's ways to get around that through encryptions, zero knowledge proofs, stuff like that. So you need to have audit trails for ai, and you need to be able to have them open so that people can look and say, is this what want? Because sometimes the harms emerge over time. A lot of things that are wrong with social media are not things you could see in a day or two. They're things that occur when you have it running for a long time in a changeable situation. So you need to be able to have audit trails that go back that far. And that's really what the suggestion is, is look, we don't know what regulations we need and what we don't. We know that in things like say electricity, we don't have 37 ways of how you have to use electricity.

39:11   We just know that electrocuting people are causing fires or bad, and we're going to approve designs, not law, but regulation standards that keep that from happening. So for instance, on every electrical cord in the United States, there's a little UL tag that's a cooperative, a not-for-profit cooperative that tests things and has standards. And if a company tests their thing and meets those standards, then they don't get sued as often and they don't get a black eye in the press. That works. That works everywhere. And we need to begin that way until we find out things that are particularly things that AI does that are especially bad, and we want to be able to detect those quickly and jump on. So that's the suggestion.

Jon Krohn: 40:05   So it might not be the most surprising thing if not too far in the future. The big Frontier labs agree to have the same kind of regulation as that UL electrical example that you had where we can kind of trust AI systems to a certain extent and not burn the house down.

Alex Pentland: 40:22   Well, so that's already beginning to happen. There's a California law that got passed that says that the big guys have to reveal what they do for AI safety. They have to publish that openly. And so now it's pretty interesting to go and compare who's a little more cautious than whom. And now we're going to be sort of asking, this is part of the stuff that we're doing at Stanford with consumer Reports, which ones are really sort of in your court and helping you and which ones are just fleecing you? You'd like to know that? And the fact that they have their safety things, you can begin to correlate that with what they do and figure out which safety things work and which are just hot air.

Jon Krohn: 41:05   Right. All right. Thank you for summarizing the things that are happening today with that AI safety regulation as well. So we've talked a lot about data regulation, data governance. We've also talked about your academic work, your book. I would like to now with the final chapter of this podcast episode to talk about your entrepreneurship. So your entrepreneurship program has spun off more than 30 companies, including the largest rural healthcare service delivery system in the world, secure privacy, preserving payment solutions, a popular delivery route optimization platform, and a biotech startup that reduces livestock methane emissions with microbiome interventions. So some amazing enterprises or startups there, companies there that all have a positive social impact. So I don't know if you have general advice for the audience, maybe lessons learned about scaling innovation responsibly, ensuring that beneficial social

impact isn't just a byproduct, but a design principle from the get-go.

Alex Pentland: 42:10   Well, I think there's probably two things that I would sort of recommend. One is you need to look where other people aren't looking. And most people are pretty shortsighted. They're looking to make a buck and that's about it. But a sustainable buck is a more fundamental problem that nobody's really doing anything with the healthcare thing, we put simple ais on phones before there was wifi and now nurses and doctors and things like that could begin serving people and helping pregnant women and stuff like that. People just weren't thinking about it. They said, oh, that's too high tech. We can't bring that into these poor countries. Well, actually things get better pretty fast. So that's the other principle for a given level of performance, the cost of AI haves every three and a half months. Think about that. That makes Moore's Law look like the slowest thing in the world. It's 10 times the performance per cost each year. It's like, oh my god. And so you can begin doing things that are sound really crazy, but by the end of the year it won't look so crazy in two years from now, it'll look like. Why didn't I think about that?

Jon Krohn: 43:32   I really appreciate that Entrepreneurial advice related to your entrepreneurial advice is maybe some of your research that bridges both academia as well as having potentially lots of real world commercial applications. You helped pioneer the concept of the living laboratory, designing research that unfolds within real world human contexts. Tell us more about that. And as kind of a follow on question, once you've told us about these living labs, tell us about how it might evolve with new technologies like digital twins and synthetic data becoming so capable following the same kind of, I don't know if digital twins, actually they must be things like digital twins in synthetic data would be following that same kind of 10 x

performance per cost capability that we're seeing with LLMs in general. So yeah, tell us about Living Labs and how that 10 x multiplier per year is impacting them.

Alex Pentland: 44:32   I mean, again, go back to sort of the social media thing. Social media was supposed to be the greatest thing ever when it first got released. Oh, we're all going to be have a voice, isn't that wonderful? And then it turned out over time, things were a little darker than we thought. And the only way you can really tell that other than people sort of finding little data to look at is to have a group of people who will agree to live in the future. So you could have a town that says, look, we'd like to try out AI transportation systems. What do you think? And if everybody says, sure, let's give it a shot, then you get to see how well the AI transportation system does in a real world thing and you get to see the unanticipated effects. Those are almost always the ones that you care about as the things you didn't think of.

45:25   It turned out to be significant. It's one of the things that we're doing with the Consumer Reports Loyal agent project is trying to ask, well, what does this AI agent stuff do that we aren't thinking about and really ought to be doing something about now? So it's having people live in the future, giving them this sort of state of the art or beyond state-of-the-art sort of tools to do. And so digital twins are part of that because what you're really trying to do is understand human behavior with this new component like AI agents. And so you'd like to be able to try different things, but you can't try everything with the people because they've got to get on living their lives. So you build a digital twin, so you build something that says, well, this is what each of these people or these people in aggregate, this is how they behave.

46:20   This is a good model of what we've seen. And you try to validate that. It's a predictive model of what you will see.

So now you can look at different designs and say, well, what are people going to do with this? Are they going to run off the end and drown in the pool or what is it going to be? Great? And the way you do that is you build a model of the people digital twin, the fact that digital twins are possible and not just possible. That's really where we're going when we talk about these things. The compute is dropping dramatically in cost. It's hard to believe how fast that's dropping data. There's more and more data. The big question about the digital twins and the living lab is how do you keep the data safe? How do you make sure that when you do these experiments, you don't do anything really bad? And digital twins help a lot with that, right? Because they aggregate the data in a way that can preserve privacy, for instance. And if the digital twin does something really stupid, at least you haven't killed somebody, right? You better go back, rethink what you're doing, but it's a way to avoid the investment and the harms.

Jon Krohn:	47:40	Alright, fascinating view on digital twins there as well. Sandy, thank you so much for that. This whole episode has been sensational. Really appreciate your insights across society, academia, entrepreneurship. We've benefited so much from your rich experience throughout this episode. Before I let my guests escape, I always ask for a book recommendation. Do you have anything for us other than your own book? Which of course all mentioned again now, shared Wisdom new release. If you enjoy today's episode, I'm sure you'd love picking up a copy of Shared Wisdom. Sandy, what else do you have for us?
Alex Pentland:	48:16	Well, I think a book that I've seen just recently that I think is really interesting is a thing called Flash Teams, and it's Valentine and Brookstein, I think it is. But basically it's how you can build new products using these AI tools really, really fast. I don't know that this is the technical book that you want, but it gives you examples

and sort of stretches your minds in ways that I've been alluding to, but it gives a lot more in the way of examples.

Jon Krohn: 48:52   Excellent. Thanks, Sandy. Yeah, it doesn't always need to be a technical recommendation, and we actually get a lot of novels recommended on the show, and so yeah, it's your choice. Appreciate that one there from you. And final question for you is just how folks after this episode can be following you and getting your thoughts. Should they follow you on social media? And if so, which platforms?

Alex Pentland: 49:18   Yeah, certainly. You can follow me. I post sort of news of what we're doing on Facebook on X and LinkedIn, and so any of those work, you can just sort of Google me and see. They're usually Alex Pentland with either underscores or dots or something like that. Right.

Jon Krohn: 49:42   Well, also we should have them in the show notes for folks as well so that they can find it very easily. Thank you so much, Sandy. Really appreciate having you on the show. And maybe we can check in again in a few years and see how research on your side is coming along.

Alex Pentland: 49:58   I'd love to. Yeah, that's great. Thank you very much guys. Good questions. Enjoy the conversation.

Jon Krohn: 50:05   I'm lucky to have an amazing researcher, someone named Serg Masís, who's a great data scientist in his own right, very, very kindly, spends a lot of time getting great research done. Yeah,

Alex Pentland: 50:19   Yeah. Okay, good, good.

Jon Krohn: 50:21   Exactly. Alright, yeah. Thanks again and catch you soon.

Alex Pentland: 50:25   Good. Take care.

Jon Krohn: 50:29   What an honor to have professor Alex Sandy Pentland on the Super Data Science Podcast. In today's episode, he

covered how AI systems fail when they treat humans as logical engines rather than social foragers who learn by watching others. He talked about the parallels between Soviet Central planning through early AI and today's systems, why Washington DC residents wanted AI to defend them against city bureaucracy rather than optimize city services revealing what real people actually need from AI data unions as the next evolution after labor unions and agricultural co-ops and living labs and digital twins as tools for testing AI systems in real world contexts before deployment. That's it. 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 Alex, Sandy Pentland social media profiles, as well as my own at superdatascience.com/949.

51:25   Thanks of course to everyone on the SuperDataScience podcast team, our podcast manager, Sonja Brajovic, media editor, Mario Pombo, partnerships manager, Natalie Ziajski, researcher Serg Masís, writer Dr. Zara Karschay, and our founder Kirill Eremenko. Thanks to all of them for producing another super episode for us today for enabling that super team to create this free podcast for you. We are deeply grateful to our sponsors. You can support the show by checking out our sponsors links, which you can find in the show notes. And if you ever want to sponsor the episode, sponsor the podcast yourself. You can get the details on how at john c crone.com/podcast. Otherwise, help us out by sharing the episode with folks who would like to listen to it, review it on your favorite podcasting app or on YouTube, subscribe if you're not a subscriber. 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. Till 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.