



SuperDataScience

**SDS PODCAST
EPISODE 1037:
RECURSIVE
SELF-IMPROVEMENT:
THE EUREKA
MACHINE, WITH DR.
RICHARD SOCHER**



- Jon Krohn: 00:00:00 My guest today thinks AI that improves itself with no human researchers in the loop, that is recursive self-improvement, could arrive in as little as a year and he isn't worried about it. Welcome to episode number 1037 of the Super Data Science Podcast. I'm your host, Jon Krohn. Today's guest is the legendary Dr. Richard Socher, CEO and co-founder of Recursive, a startup that came out of stealth this year and has already raised over half a billion dollars to build recursively self-improving super intelligence that automates scientific discovery. Richard is also co-founder and CEO of the AI search engine, you.com, and a founder of the AI-focused venture capital firm, AIX Ventures. He was previously chief scientist at Salesforce and his brand new book, The Eureka Machine is already a national bestseller. In this episode, Richard explains how AI could deliver a century's worth of scientific breakthroughs in a decade, why fears of an overnight AI takeoff ignore physical constraints, and which software and jobs will last no matter how capable AI becomes.
- 00:01:05 Enjoy this one. Episodes don't get better than this. This episode of Super Data Science is made possible by Anthropic, Dell, Nvidia, Origin, and Palo Alto Networks. Richard, welcome to the Super Data Science Podcast. It's a treat for me to have you on the show. How you doing today?
- Richard Socher: 00:01:22 Great. Glad to be here. Very exciting to chat with you.
- Jon Krohn: 00:01:26 Yeah, I've been following your work closely since 2017. I was running a deep learning study group in person in New York, and we used your CS224D course as an introduction to deep learning for natural language processing. We learned all about glove back then from you. Those were really great pre-pandemic in-person learning times for me.



Richard Socher: 00:01:48 Yeah, you're not alone. The hugging face founders started a study group also in New York and that worked out well for everyone. And yeah, it's the OG first class in the world on neural nets for NLP.

Jon Krohn: 00:02:02 Yeah, I wish I'd been able to turn a New York deep learning study group into a billion dollar exit to Nvidia, but maybe I still have the threads going somehow that we'll lead there someday. Well, so let's talk about what you've been up to more recently. So you're the CEO and co-founder of Recursive. You co-founded that business last year. It came out of stealth earlier this year. It's already raised over half a billion dollars. And the idea behind Recursive is that you build Recursive, self-improving, super intelligence to automate knowledge discovery. Do you want to tell us about that business?

Richard Socher: 00:02:40 I'd be happy to, yeah. It's exactly what you said. We believe that after seeing natural language processing and AI hit several massive inflection points because we replaced human processes, expert designed features with bitter lessened, larger scale automated processes, we basically realized, well, we've done this with feature engineering, then the community switched to architecture engineering. Then I had this crazy idea in 2016 and 18 to try to replace all that architecture engineering with one neural network that you can just prompt with any kind of question. It's called DECA NLP, that paper famously rejected, but cited a handful of times by the early GPT papers from OpenAI is kind of putting forward this idea that really you can just have one model and prompt it with any kind of question and always get an answer back. And so that led to unifying all this architecture engineering and now letting the AI figure out how to deal with different kinds of questions.

00:03:49 And now I feel like the field is stuck again, but at the next meta level, and that is basically manually doing research, manually coming up with ideas, then mostly using AI now



also to implement more and more of those ideas and then validating those ideas, which is also non-trivial with a lot of compute and so on. And we figured out let's try to essentially automate one more step and that might be the last necessary step for all of AI. And so that is when you automate the scientific method and you apply that automation to AI itself, then before you know it, you get to recursive self-improvement. And a lot of people use that term very loosely by when you see blog posts from OpenAI say we're getting better at self-improvement because look, we're using codex more and more. But that's still, I would say, argue, weak form of self-improvement where you still have humans majorly in the loop, human experts major in the loop.

00:04:51 And that makes sense when you have thousands of them on your payroll and you pay them tens of millions of dollars every year to do this. And so we think that it can be done better and differently and fully in a strongest form recursively self-improving. Of course, even there, humans will set the goals, set the rewards, set the environments and so on, but inside those AI should really come up with its own ideas, implement them and validate them.

Jon Krohn: 00:05:20 A lot of people use RSI as a scary term, but it's in your business name. For you, it's obviously not a scary term.

Richard Socher: 00:05:28 That's exactly right. Yeah. I think it's really fascinating to dissect the various groups that have non-zero P doom, and it's very easy for the human imagination to come up with scary scenarios. I think you can, in various levels, try to put science behind them. I think at the same time, we've seen that many times before. The number of experts that predicted financial crises correctly, the hundred out of the five financial crises are correctly predicted as a joke. The overpopulation crisis was huge and people were worried there'll be so many people, too many people. The amount of crises and fear around nuclear energy, around



GMOs, there's so many fears that people have had that ultimately held back humanity. The number of people who died from keeping coal around for longer is so much higher every year than all of nuclear, including Chernobyl and Fukushima and other disasters combined.

00:06:36 And so we have to be careful a little bit and not dismiss progress. Obviously, AI is omni-use technology. I'm not saying they're zero risk. Maybe chatbots should have ratings like movies. Maybe we should have FDA, controlled drugs, which we already do, and transportation authorities control self-driving cars, which they already do. And think about a lot of the applications of AI and where they're touching human lives, we should be careful. But I think a lot of people have a very vivid imagination when it comes to the apocalypse.

Jon Krohn: 00:07:06 Yeah, I love the analogies that you're making there related to coal power, nuclear power. Looking ahead, it seems like we're on the cusp in the next few years of having commercial nuclear fusion, for example, and AI, especially an AI system that is helping bring scientific breakthroughs, can make a big difference in realizing nuclear fusion faster, other clean energy technologies, storage technologies. And so there's a lot of resistance, particularly in the United States today, but in probably lots of countries all over the world around building data centers, empowering data centers. And it seems like it's kind of a similar argument where it's if we can accelerate getting to fusion energy faster by building an AI data center, to me, that seems like the way to go.

Richard Socher: 00:07:53 That's right. I think a lot of folks, and this is partially because of the media, partially just human sort of priors of being more fearful than joyful about the future. There's so many positive things and that's kind of what my book is about, the Eureka machine, where I basically lay out how impactful AI will be to scientific discovery in physics for energy generation, in chemistry, for energy storage, in



biology for creating healthier people, and really being able to eradicate disease, which sounds as impossible as it sounded 10, 20 years ago to automate all of natural language processing in one model. But if you have enough data, it'll be possible to understand our brains better of neuroscience, to improve our economies with AI economists, and to really go even beyond earth into astrophysics and understanding that better. The book lays out where all of that is already happening with AI, where it's already seeing progress.

00:08:49 But then when you combine all of these things, you can create this ultimate eureka machine. And I wish in the grand scheme of things, data centers have helped a lot of communities. And of course you have to be careful, especially depending on how they're powered, where that exhaust goes, and you have to be conscientious around where exactly they're put and how loud they are. But these are all fairly fixable issues. There's this whole lie about them using up all this water as if they're just steaming that water back into the atmosphere. Almost all data centers these days are closed water systems. So there's a lot of fear and uncertainty and doubt that is spread. And I'm always not quite sure whether it's a political SYOP or if it's just people are worried about things and you should actually talk to all the communities that are there and hope that they can be responsive to the facts, which are mostly positive.

Jon Krohn: 00:09:48 Yeah. Well, you've certainly got an ally. I guess it shouldn't be surprising that maybe somebody who hosts a podcast like this is very positive about the future and where this technology can take us. Everything that you covered in your book is brilliant. To give a bit more context on that, I'm glad we're segueing to the book. It was published last month at the time of this episode coming out. So brand new book called the Eureka Machine. And in it, you argue that AI can connect every stage of the scientific process from hypothesis generation



and simulation to experimentation and validation to potentially deliver a century's worth of breakthroughs in a decade. And I finished doing a PhD in neuroscience, but focused on AI applications in neuroscience, finished that in 2012. And if you had asked me in 2012, if we would have the AI capabilities that we have today, at any point in my lifetime in 2012, I would've said, "That's crazy.

00:10:51 There's no way we can have machines doing the things that we have them doing in 2026." So pretty mind-boggling, and that kind of perspective does make me think that a century's worth of breakthroughs in the coming decade is not an unreasonable thing to say. So tell us a bit more about your book and why you wrote it. And also congratulations. I just found out this morning that it is now a national bestseller.

Richard Socher: 00:11:19 It did just become a national bestseller, which is really exciting and I'm really happy to see that. I think in the book, like you said, we lay out these different things and different fields and how AI will impact them in very positive ways. And I think a lot of people agree that science is not a zero-sum game. Oh, some person gains more, that means other people will gain less. Science really creates net new value and new capabilities, and many times also new jobs that follow from those innovations. And so I think it's a really important part of the narrative. It almost feels like the future needs better marketing these days. And indeed, when you look into each of these fields like biology or what calculus did for physics, I think AI will do for biology. It's the new and right kind of language to not just assume we understand the small pieces, which we got pretty good at in science, but actually understand the very complex networks that put them together.

00:12:18 We understand each neuron, but we don't understand how the whole brain works. We understand most of the specific bacteria in our gut, but we don't understand how



the whole microbiome really interacts with different kinds of food and how that connects to your brain and maybe why some people get hungrier than others after they had a lot of sugar. Just all these subtle things and sometimes very profound things that we don't understand because we don't understand the complexity of the system. So we have to weave all these separate parts back together and AI is the perfect technology for that. And we can indeed look at what had happened with natural language processing where with enough data, and that is kind of the famous bitter lesson, with enough data and a large enough neural network, we can actually learn a lot of things and ultimately have a full conversation.

00:13:09 Now, currently, biologists interact with biology as a natural science. It's not yet in the eyes of many programmable science, but that's what AI will enable us to do with progen and many other AI models. But when that is sort of akin to how linguists had studied natural language where it's like, well, I know mutation, so not this or didn't think that. And I write these regular expressions because I'm a linguist and expert in them and I study Chomsky and Xbar schemes and syntax trees and so on. And at some point you can throw out most of that and just use a very large neural net and that will do better than any of that expertise. As we get enough data, that kind of insight will be possible in biology too and in other sciences.

Jon Krohn: 00:14:01 Yeah. The bitter lesson that you alluded to there is a famous blog post by Richard Sutton from 2019. I'll have it for people in the show notes, but yeah, as you're alluding to there, it's basically this idea that human expertise and trying to fiddle with the features in a model should be and how to get some output from a machine system based on some inputs. It's the bitter lesson for Richard Sutton is that - It's bitter because



- Richard Socher: 00:14:26 It's the expertise. So many papers, like 90% of all papers in AI were about people being really clever about this hack and this insight and this subtle probability distribution and priors and sampling cleverness and so on. And it's just a large neural net, simple gradient descent. And there are a lot of obviously complexities as you scale those up and be more clever about how to optimize them and so on. But the simpler the methods are, the more likely they will work at larger and larger scales.
- Jon Krohn: 00:14:58 Yeah. Those scaling challenges are mostly about engineering more efficiently than having some kind of scientific new idea.
- Richard Socher: 00:15:05 That's exactly right. And I think ultimately the most amazing end state for most sciences is actually to become an engineering science, which is also maybe counterintuitive. But once you understand the science of the atom, you can build nuclear power plants. Once you understand the protein sequences and how they are formed and created and how they're impacting the body, then you can start to engineer them more and more. And so that transition is something that is to some people it's sad, but if you truly care about the outcomes, the impact that your scientific field can have, you're mostly excited about that transition.
- Jon Krohn: 00:15:47 Yeah. Hopefully someday the neuroscience that I studied, it's difficult for me to even conceive of, but to have neuroengineering as the standard as opposed to neuroscience, tremendous potential there, a complete reframe of the way that we think about studying the brain. So in your book, you describe alien hypotheses as in this new life form or new cognitive capability that feels alien. And I think that comes from this idea of going back to, was it go or chess? I think it was with Go, with Alpha Go talking about this idea of an alien intelligence and playing against, because there were strategies that Go



experts had never seen humans play before. And so these kind of alien hypotheses that AI systems come up with and even the hallucinations that they have, you argue in the eureka machine that these are a feature, not a bug. And that's part of what will allow AI to bring about so much scientific breakthrough.

Richard Socher: 00:16:51 That's exactly right. AI helps you basically explore this vast combinatorial space. The combinatorics are insane if you try to think about all the ways you can cue words together. And it's such a massive, unfathomably large space of all possible sentences and thoughts and combinations thereof. But with AI, we have a really good prior to take what has been of the past and then explore reasonably different ideas that are interestingly different and then think about recombining those as well. And so sometimes of course hallucinations are a feature, but sometimes they're also a bug. If I just ask a factual question, then we should just rely on the facts on the ground. And that's why part of the Eureka machine requires this kind of search engine of existing human knowledge, which is why I still love running and working at you.com where we build a web search engine for agents that actually can be accurate, up to date and have citations for the facts that come out of a conversation.

00:17:57 So I think that is going to continue to be a crucial part of the entire stack. And I think it's an interesting example also of where the world might be going in terms of a lot of the technologies we've built for humans. They can be re-imagined as infrastructure layers for AI. We had the browser for humans, we had the GPUs for human gaming, we had the internet of course, and we had search engines for humans to choose which link to click on and so on. Now we can reimagine what those technologies would look like if agents are using them as tools when we give them tasks.



- Jon Krohn: 00:18:36 Yeah. It's interesting that you mentioned you.com in your answer because I was going to bring that up in my next question related to the whole Eureka machine thing because you patented work at you.com. So to give the audience a bit of context, you co-founded in 2020 with Brian McCann who was on the show two years ago back in episode 835 if people want to hear lots about you.com from Brian there. But you're still co-founder and CEO of you.com today on top of Recursive. And at you.com, you patented work on search that suggests trustworthy AI needs providence and access to external knowledge while the Eureka machine must eventually discover things that no source can yet confirm. It's a very different kind of world I guess that Recursive is trying to bring about with the Eureka machine. So if an AI discovers a scientific insight, what's the minimum kind of explanation package that we need?
- 00:19:37 A mathematical formula, peer reviewed paper, a model tested in a simulated environment, maybe even running a physical wet lab with a machine. How do we bridge where we are today where it seems like we're very rarely getting kind of new information out of our AI systems to getting new information and verifying that information in the future?
- Richard Socher: 00:19:59 So yeah, we have to have these four pillars of the Eureka machine where you have world knowledge, you have scientific measurements, you have simulations, and then you have real life experiments. And I think on top of all of that, we'll have this agent swarm. And then indeed there might be some explanations that are not very satisfying because they're so complex, we cannot have the intuition anymore of our somewhat puny little brains to understand how these trillions of parameters after using many different layers of non-linearities and attention mechanisms and activation functions and large matrix multiplications, how they actually predict the next word. But of course we understand it sort of intellectually on



some level of abstraction. And so the question is, are we going to be okay probing this model, trying to understand it better, even though initially we might not have full access to a deep intuitive understanding of it?

00:21:01 And it's a little bit similar to if I asked you, why did you move this strain of your pinky finger when you're trying to move your steering wheel to the right to stay in the lane, you don't have conscious access to that. You don't really have conscious access to your early visual cortex of why did you think that was a dot versus a gradient of this angle and things like that, that neurons that we know fire in the visual cortex for certain visual stimuli. And so I think we may have to make peace with some of that as we progress in building more and more complex models of reality.

Jon Krohn: 00:21:39 Yeah. It brings up interesting thoughts around how do we, if we can't understand intuitively a machine's theory or explanation or recommendation on something that we should be doing, assuming that there's some kind of decision points of human in the loop where we say, okay, that was a great idea or we trust that idea even if we don't know intuitively that it's a great idea. And so we're going to go ahead, we're going to build that machine that that machine asked us to or take this action. If there's humans in the loop on whether the action happens or not, and we can't understand why we should move forward or not move forward with a particular action item, what do you think the minimum amount of information should be like a model tested in a simulated environment or something like that maybe? So

Richard Socher: 00:22:41 Yeah, I mean, I think we will have various levels of understanding. Sometimes we have to acknowledge even for some standard medications that everyone uses, they thought GLP-1s were working in this way. And then later on they found actually it was all about the brain and how the brain thought about things, not about how the



stomach though it was full and whatnot. So there are all kinds of theories and we have to, of course, probe them every time and improve them over time. All models are wrong, but some are useful and will improve their precision over time. So I think that is a very common thing in science. And of course, when there is an idea that is extremely powerful and has a lot of ramifications, then it is much easier to probe a neural network, to probe a complex model than it is to probe the natural world.

00:23:26 In some sense, the models become a substrate of intelligence and then that substrate is much easier to evaluate than it is to really probe the brain where we can either have a global understanding of how oxygen is used in FMI scans, but that's really poor temporal resolution or we have great temporal resolution, but there's only for a few neurons that where we literally stuck an electrode into the brain and then measure whether that neuron fires or not and that ignores all the chemistry. And so we don't have a really good way to truly measure what happens in the brain at high temporal and spatial resolution simultaneously. So that will be much easier to do in neural networks. And yeah, I think having certain inventions go through the alignment with existing knowledge, scientific measurements, simulations, see how they perform under many different circumstances, and then experiment in the real world carefully.

00:24:25 Those are the four sort of steps that I outline and that I think this automated swarm and eureka machine will be able to do so much more efficiently than human scientific communities.

Jon Krohn: 00:24:40 Makes a lot of sense. To dig into this a little bit deeper, two years ago you predicted that AI designed proteins would have a ChatGPT moment within the next five years. And so at the same time, you noted that biology advances more slowly because experiments happen in the physical world in biology. And you gave examples there of us using



medication that we don't understand how it works. And so for me with a neuroscience background, psychiatric medications, we basically have no idea how they work. I don't think there's a single antidepressant that we understand how they work. Even general anesthesiology, which is one of the oldest types of medicine that we still use today, I don't think there's any understanding of how anesthetics cause unconsciousness. Even this idea of how consciousness arises from the brain, there aren't clear understandings yet today. So in the book, you call high fidelity simulation the safest and most powerful arena for AI exploration, yet models of brains and other complex systems like ecosystems, climates, those are inevitably incomplete.

00:25:48 So how do we prevent the safety sandbox itself from becoming a source of false confidence when simulated discoveries move into the physical world?

Richard Socher: 00:25:57 It's a tough one. I don't think humanity has the full answers. We do know that we can use neural networks to simulate physical environments better. For instance, weather prediction, a company called Atmo that just was recently acquired and I was an investor in. They showed that you can, instead of modeling everything from raw physics fundamentals, you can actually use neural networks to approximate parts of those models. And that is very just much more efficient and faster to run. It used to be a few years ago that most of the computation went into weather modeling actually, and I think that has changed now. So we can build higher and higher fidelity simulations and it's surprising even for people. I mean, there are companies now that simulate a large user group and you can prompt the same LM with different ways like you're a 45 Midwesterner, blah, blah, blah, and you're a hip New Yorker and from the Bronx.

00:26:57 And then now how would you react to this thing? And interestingly enough, statistically speaking, the outcomes



are more and more similar to what you then would see with real life experiments. And so I think we can, again, acknowledge that all models are wrong, but some are useful and we can make them more and more useful by scaling them up and increasing their fidelity. And I think one such example that I talked about in the book and we had written a paper about, it's called the AI Economist, where you can make certain assumptions about how much people want to work, what their utility functions are in terms of at some point, I don't want to work 14-hour work weeks, work hours every day. Some people may say, I do often, and then you make different assumptions around lots of agents and then you look at if they're mostly acting in their own self-interest, how would you tax and subsidize that community in order to achieve a goal that that community would like to achieve, maybe increase overall productivity, multiply it with equality, or maybe add sustainability, whatever you want to do to your economy, you can add those requirements and rewards.

00:28:05 And then you have this two level reinforcement learning system where you have an AI economist, and that was the name of the paper back during our Salesforce days. The AI economist suggests taxes and subsidies and all the individual agents are just optimizing and maximizing their own utilities, try to block each other from resources, try to pay as few taxes as possible, try to do clever sort of tax evasion and whatever. And you basically can now iterate billions of years of taxation and subsidization and see what is the actual best outcome. Now, if you can have a simulation that's good enough for that, then we could avoid or could have avoided as humanity, many different wars where people fought over to some degree such ideas and ideologies. And I think that would be also very useful to have, at least as an advisor, have super intelligence as an advisor and see if certain goals that you have are really truly achieved with the instruments, financial instruments that you propose.



- Jon Krohn: 00:29:07 In your book, this chapter on economics is chapter five, you titled it How to Make Economics a Less Dismal Science. And I guess tying back to something we were talking about earlier in the conversation, I guess it's about making economics and engineering discipline as well.
- Richard Socher: 00:29:20 That's exactly right. And I think right now, if you look at economics, the whole field is very just ideological. And when they do have some math, it's this beautiful math that shows that it's provably the right taxation scheme in a one-step economy where people make one economic decision and then no decision ever again. There's no adaptation, there's no subjectivity and maybe suboptimal egotism or an adaptation to different environments. It's just a one-step economy. That's where the field of economics is still stuck in. And we submitted this paper and we had just some economic ethicist or something that's reject the paper from science and nature. That field has not yet had not even its ChatGPT sort of GPT-3 moment, but it hasn't even had its machine learning AI moment of any kind yet because it is mostly ideological. If you want to work in that department, you have to have those ideologies and then the math will work out such that those ideologies will work for you.
- 00:30:20 And I think it would be really great if we try to make those policy recommendations a little bit more objectively in the future.
- Jon Krohn: 00:30:28 Yeah. Going back almost two decades in your career now, you were one of the developers of the ImageNet dataset with Fei Fei Li. And so this ImageNet dataset, probably most of my listeners are already aware of it, but that is key. Having a very large number, millions of labeled cases of images across thousands of different categories was critical to allowing breakthroughs like AlexNet in the ImageNet visual competition. And it allowed this bitter lesson kind of effect where instead of trying to figure out,



okay, a cat should have two ears and a mouth and programming all that and trying to have feature detection, just allowing data and compute using the ImageNet dataset with a sufficiently large deep learning model to automatically identify patterns was the route to success. And you've now seen that pattern many times over your career where data, a sufficiently large neural network are capable of getting us approximations where we don't need to understand in the weather examples you were giving, all the fundamental physics underlying it.

00:31:35 In the economics examples you're giving, we don't need to understand... Well, I don't think in economics they're even trying to model down to individual.

Richard Socher: 00:31:44 They're making extremely oversimplifying assumptions. The economy is a linear system, which is totally nonsensical.

Jon Krohn: 00:31:50 But yeah, it seems like across disciplines, that same kind of machine vision breakthrough that your ImageNet data set allowed for is as we gather the data, as we have self-improving neural network systems that work with those data, help label those data, help create those data, I suppose essentially across any imaginable discipline, we can be turning from a science to an engineering practice.

Richard Socher: 00:32:21 That's exactly right. And I think it'll be particularly powerful as we make biology into a programmable science because that is very close. Most people don't care about the field of medicine until they're sick and then it's all that matters. And so I think as you've gone through more and more serious health scares, as we all age and aging sucks and basically it's sort of a disease eventually too, we will become more and more clear that AI will be saving grace and especially now with populations going down, which is also a big threat for humanity, having fewer and fewer people to hold conscious and intelligent life up. I



think longevity will be really an important and fun application area of AI.

- Jon Krohn: 00:33:09 For sure. Perhaps we could even simply consider aging to be another disease that needs to be cured.
- Richard Socher: 00:33:15 It feels weird and it's certainly pushing the overdo window of acceptable discourse for a lot of people right now, but it's kind of similar to the pill, reproduction, reproductive rights and the ability to have control over that felt really unnatural and weird to a lot of people. But once you're aging, you're like, "Oh, I wish my knees didn't hurt. I wish I could still work harder or run further or go on longer hikes, but I can't now and my life is shrinking more and more." Some people will be okay with that and let them die of old age if they want to, but more and more people will be like, "It would actually be nice if I could still do these things I used to be able to do." And so I think as you get older and as you have a lot of that agency and will to live, I really hope that AI will be a tool for us to let people accomplish that.
- Jon Krohn: 00:34:08 Yeah, it sounds pretty good to me, of course. And yeah, it's the kind of thing I'm like, wow, I wish we could have figured out that faster for older relatives that have already aged or passed on. Obviously, I don't need to convince you that it's an unbelievably exciting time to be alive and some things that have been unimaginable for all of the rest of the history of living things and humans might be happening in our lifetime. So specifically related to that, so I was talking about how chapter five was about economics. Chapter six is your neuroscience chapter. And in it, you talk about something that I've never talked about on the show. And actually I know very little about myself, which is brain organoid development. Richard, what are brain organoids?
- Richard Socher: 00:34:55 Yeah, so organoids are tiny little mini organs with a few thousands to hundreds of thousands of cells, sometimes



a few million, that essentially mimic parts of a real organ. And there's some really cool companies like Pearlbio that use this to essentially test how drugs would interact with the human immune system or the human neural system. And they essentially showed that testing on these tiny organoids is more predictive of how a drug would interact with the human body than mouse studies. The joke is in biology, often we cured pretty much all diseases in mice. Turns out mice are very different to people. Making a mouse live 30% longer does not mean a human will live 30% longer. Curing an immune disease in a mouse or cancer in a mouse does not mean you can cure that same disease in a human. But it turns out that Pearlbio, for instance, got full FDA approval to skip animal trials for immune therapy toxicity testing, for example.

00:36:06 And that not only saves in the future millions of animal lives that are only bred to then be tested upon and dissected and used as lab animals, but you can also get drugs into the market faster, which is amazing for pharma companies because they stay longer in the sort of patented window that way because the clock started ticking earlier during the development already. And so yeah, Parabio is one such example. And yes, the brain ones actually, I also asked them, "But you're making sure those brains are very, very small." And they're like, "Yeah, it's less than a bee. So if you're not worried too much about the life of a bee, you don't have to worry too much about the life of an organoid of sorts." But it is very powerful to understand, especially the brain where we have so little understanding of even how a few thousand neurons combine and change their activations.

Jon Krohn: 00:37:08 It's pretty interesting. So something that I think about a lot, I eat meat most meals, but I actually think it's pretty... I think it's something that my grandkids, when they find out that we were slaughtering animals and eating them, I think they might be like, "Wow, I can't believe that you guys were doing that. You guys were like



cavemen." And so I'm hopeful that we can have very well simulated steak with the marbling and everything. In that case, you're like, "Okay, we're taking the conscious being out of the picture and we're growing the muscle on its own." And obviously it seems like there's very few ethical dilemmas associated with that, but it hadn't occurred to me that with brain organoids, you could potentially get to a place where you're simulating a brain so effectively that just as you were talking about with bees there, you would have this conscious thing that in the lab that you've kind of created.

00:38:05 It's an interesting mind experiment and I guess - You

Richard Socher: 00:38:07 Have to be careful not to make those so big that they truly could actually develop any sense of thoughts and consciousness and so on. And I asked the team and they're like, "Yeah, don't worry. They're like less neurons than a bee." Yeah.

Jon Krohn: 00:38:21 I mean, I'm kind of extrapolating into the future hypothetically things that we could be doing with an RSI system that has enough data and can be neuroengineering a whole brain in a lab. But anyway, we're a long way away from that moral dilemma. Let's talk more about recursive. So you distinguish between simply optimizing a model and building systems that can identify their own limitations, improve their research and training processes, and verify that those improvements are genuine. So recursive names both your company and its central ambition, which is systems that can recognize and remedy their own limitations. We have a quote that you think we could potentially get to RSI superintelligence within two years. And to get there, we obviously have to distinguish between, as you were talking about earlier in this episode, today's auto research from true full RSI. I don't know if there's more that you want to get into there right away or yeah.



- Richard Socher: 00:39:28 It might even be less than two years. It might be only one year. It's looking very, very promising. And at the same time I have to warn people of just because you have RSI and the functioning loop doesn't mean that you have this strong takeoff scenario that a lot of people who are fearful of AI then go jump directly into where they say, oh, because we have RSI, the next 10 minutes will then be like, it could dominate everything and they just forget about all of physical constraints and infinite robot armies that somehow have perfect supply chains and perfect resources and perfect money to infinite money to build these robot armies overnight. And so there's a lot of fears that just don't make any sense. And even for RSI, if you ask the RSI now to build more and more sophisticated things and larger and larger and more and more complex models, they have to also have compute to do all of that.
- 00:40:24 And so the bigger the ideas are, the bigger the models become, the more compute you will need to give to that intelligence warm to be able to achieve its goals and the goals you set for it. So yeah, that maybe is just an important side note. So it sounds like science fiction to have these recursive self-improving loops because people assume that then crazy things will happen right away, but there's still iteration cycles that will have to happen inside that eureka machine that works on itself.
- Jon Krohn: 00:40:55 Yeah, I guess it's in a bit of a less constrained way. It's kind of similar to talking about biology where we can't just because we have, if we get to a point where we have an effectively unlimited amount of intelligence at no cost with something like biology for sure, but even computer science, even for developing higher levels of intelligence, there's still an iterative process. There's still experimentation. It's not like intelligence itself can figure out how a biological system works if it doesn't have the information, if it can't run an experiment.



- Richard Socher: 00:41:29 Exactly. Now fortunately, of course, the experiments on a chip in pure software are much, much faster than those in biology. So they do have a huge advantage there, but you don't train a five trillion parameter model in five minutes either. So there's still complexities.
- Jon Krohn: 00:41:47 For sure. In a LinkedIn post, you recently said that the best way to expand humanities knowledge is through the scientific method after AI independently produced state-of-the-art results on three different AI tasks. So how much of your results do you plan to share with the community at Recursive, these kinds of breakthroughs, and how much of the dog food do you eat yourselves to improve RSI?
- Richard Socher: 00:42:13 Yeah, we're using it everywhere, of course, already internally. And we hope to be able to open source some of the artifacts that this Eureka machine produces for the community to understand and evaluate and inspect the outputs of it. Now, we won't give full access to the entirety of the Eureka machine. It's kind of like saying you can look into the brains and interview every one of our employees with no NDA forever. It doesn't quite make sense to give full access to that machinery, but people will be able to see the kinds of artifacts it produces and interact with those in the future. And so yeah, we're very excited about that. And indeed it is not a zero sum game to be able to create these new inventions and then put them out into the world.
- Jon Krohn: 00:43:08 For sure. I mean, I imagine a big part of what's so exciting about this for you is that this isn't a breakthrough that is just a commercial thing. It's something that is a breakthrough for all of society. But yeah, it will still be interesting commercially for sure. So given -
- Richard Socher: 00:43:27 That's also true. Yeah. I think both of these can be true at the same time. It's actually interesting and it's something I struggle with a lot, which is if you look at the history, a



lot of people, no matter where they are on the political spectrum, appreciate the sort of pinnacle of an era's outputs like the pyramids and some beautiful castles and villages or towers of pizza and so on. People appreciate art and music, but all of those things are to some degree results of inequality. And the more epic they are like the pyramids, the more inequality there was at the time. And so how do we balance that and make sure that everyone keeps improving their state of the world while we're still allowing for some of that really long-term thinking to create things that last thousands of years and maybe really propel humanity forward in science and research.

00:44:29 And so it's an interesting philosophical balance. How do we distribute resources into both of these buckets, if you will? And of course, everything is kind of more continuous, but yeah, it's a tough question of our time.

Jon Krohn: 00:44:43 For sure. Something that's kind of an interesting thought experiment that I get trapped in myself, just kind of laying in bed at night or on the bus or whatever where you think about, okay, what if everything does turn out as amazingly as we think it might in terms of abundant superintelligence that's effectively cost free. We have nuclear fusion creating far more abundant energy than we need. We're able to tackle all the things that today we are worried about, climate change, inequality, access to education, all of these things, aging, disease, all of these things. Let's say they all become kind of solved problems, engineered problems that have solutions. It's kind of interesting to think then in a world of such abundance, if anybody can just kind of think a thought and have that transform into an additional physical object that they want or discovery that they want or some way of gaining some kind of pleasure, it's difficult to imagine what that world was like with 10 billion people all just getting whatever they want whenever they want it.

00:45:54 How does that look?



- Richard Socher: 00:45:56 Yeah. There's an interesting book Down and Out in, forgot the title, but that kind of imagines that future. And I think there are a few jobs and a few things that almost by definition can never be fully abundant and that is fame and how much that fame kind of garners human attention. Human attention can also not scale before above twenty four seven. And so maybe we can sleep less and watch some stuff even more and be unhealthy and something, but there's only so many hours in the day for humans. And then fame might become more and more of a currency that is just almost orthogonal to a material abundance. And we know that people will eventually then just be jealous of the other famous people, even though they have now warm water and transportation. They can go on vacations all across the world and they can do all of these other material things.
- 00:46:59 But at the same time, I think when you look at human progress and there's some really great tweets and there's actually a website called humanprogress.org that I really like that shows how many hours do you have to work to be able to afford certain things and how much does that price change over time? And they just show that certain things like cotton, sugar, food, all kinds of food, but then also crude oil and coffee and lead and all kinds of other things basically are getting cheaper and cheaper over time in terms of how many hours do you have to work to get access to them. And so I think ultimately human progress has been incredible and people though forget these sort of long-term statistics and think maybe things are getting worse when for the most part they're actually getting better.
- Jon Krohn: 00:47:57 This website is great. I've got the trends page up at humanprogress.org/trends right now. I'll obviously have a link to that in the show notes. But yeah, if you want to get some, it's so easy with the way that social media or traditional media captures our attention with fear and anxiety to get us to watch more of their ads and scroll



through more Instagram stories. It's so easy to feel like everything is getting worse when you see charts like these around world gross product, poverty levels, resources, commodity availability, food supply, number of calories available worldwide or in Sub-Saharan Africa, you particularly see this huge spike up since 1990. And so yeah, lots of charts to dig into here to make you feel great about where we've been going and where we'll presumably continue to go.

- Richard Socher: 00:48:49 That's right. And again, AI and science will create a lot more of that kind of abundance.
- Jon Krohn: 00:49:00 Yeah, for sure. If you take the progress we've made in the past hundred years and you can make that same amount of progress in the next 10, then that's going to be even the charts will take off even more than we see here. Bringing us back a little bit to the present and the kinds of tools that we have today, one of the wild things of the past year for me as somebody who works in AI and writes code is coding agents and how unbelievably good that they've become. You said last year, we have a quote from you where you said coding is one of the most exciting applications for AI to become superhuman. And that was before, you said that before we had Opus 4.6 in February or March this year. I remember that was a big watershed moment for me in terms of being able to trust a system to be autonomously solving some computing problem that I give it.
- 00:49:56 And since then, Fable, six Astra from OpenAI, it's become easier and easier and easier. It's become so good in fact that even legendary programmers are adopting it. Linus Torvalds likens it to compiling and Ruby on Rails creator DHH shared that he no longer writes or even reads raw code by hand. So if AI generated code, Richard, it might as well be assembly that nobody reads. I wonder if there's even a point in code being human readable. Maybe it'd be



more token efficient if we just had it in embedding layers or something.

- Richard Socher: 00:50:36 Yeah, I think it's not crazy to think about different ways. I do think ultimately if something really, really matters, like there's some very important proof point for a cryptographic protocol or something where you're like, that proof better be right otherwise we cannot use this as a tool, then you do want to be able to go and still understand that code and have it be somewhat human readable, I guess. Most humans cannot actually read programming code, but at least a few humans should be able to read it. But yeah, I think it could eventually result in new programming languages. Certainly it's a new paradigm. And this is something that I think computer science has been better at as a field than almost any other field, and that is build tools and enable everyone to keep moving up in layers of abstraction. No one has to work in assembler anymore.
- 00:51:33 No one has to even work in C++ anymore and think about pointers and where that object is in RAM. I mean, it's always useful to have a few experts still know all these layers, but just in the grand scheme of things, the majority of people in computer science don't need to fully understand all those lower level transistor layers and so on. And now we've literally gotten to meet the rest of humanity as a field in terms of level of abstraction and now anyone can use the tools. Imagine we can enable that for biology and disease and healthcare and things like that. I mean, it is a matter of time if we keep pushing in the field to make that happen. But yeah, I think coding, actually I predicted this even years before, coding is a perfect environment because it's simulatable and verifiable and anything that has those two properties, AI will get to superhuman capabilities in.
- 00:52:26 And so a lot of interesting questions will be like, how can we pretend to make things simulatable or verifiable over



time in other fields? But for programming, it was fairly clear to me it's less simple than of course playing games where you see everything you need to see on the board. And that was even more obvious to me that AI will solve it. But yeah, programming continues to be incredible. And we kind of joke the famous saying of, I think Mark Andresen, software's eating the world. Well, AI is eating software. And now humans can really guide all of that. And at the same time, I think programming is a way of thinking also. And we know models get better when you ask them to learn how to code. They get more logical, they get better at reasoning, chain of thought and so on. And so I think in that sense, just like people still study ancient Greek texts, not because they make more money in that way, but it is sometimes an interesting way of thinking about the world.

00:53:30 I think similarly, coding could be one of these stations of the gyms for the mind that we need to build where you could go to the gym and have your robot lift the weights up and down, but you just don't get the gains. And so people see those gains visually when they go to the gym and they like it. I think we need to try to allow them to see that gains for the mind are just as important and your brain will wither away just like your muscles with away if you don't use them. And so to have an interesting conversation, to have a rational argument with others, it is helpful to know still how to code and generally how to think.

Jon Krohn: 00:54:08 Yeah. I think that mind gym idea makes a huge amount of sense. I think probably a lot of people on the planet wouldn't go for it voluntarily, just like they don't voluntarily go to lift weights. And it's an interesting thing because it seems like most people want to look fit, want to be fitter. And it's like a color by numbers kind of thing. It's not very complicated. You just got to get your gym clothes on and go there and you can have a pretty simple program or go to some classes and yet most people don't



do it. And I think this mind gym idea is probably similar where it's probably going to be a minority of people that opt for it in a future where we don't need to be working or don't need to be using much cognition. There will probably be some of us who nevertheless still want to be doing math problems for fun, doing programming for fun and doing those mind gym exercises.

Richard Socher: 00:55:00 That's right. I think that's exactly right.

Jon Krohn: 00:55:02 The last thought, so I've got some audience questions for you based on a LinkedIn post that I recently asked. But just before we get there, the last thing that I want to end on is this constructive optimism philosophy that you've talked about. So I feel like it ties together a lot of what we've been discussing in this episode. In thought leadership and engagements that you've had with the World Economic Forum, with Ted, with the Aspen Institute, you've advocated for a constructively optimistic framework for guiding superintelligence. And so you've argued, for example, and in this episode and in those other forums that the future needs better marketing and to expect a new scientific renaissance. So in order to get there, it sounds like your work highlights some of the critical things that we need to get there. So precise reward engineering, which isn't something we've talked about in the episode, but is something you've been talking about a fair bit recently, robust stopping conditions, responsible power distribution to ensure that advanced systems remain aligned with human agency and the meaning that we wanted to have.

00:56:02 Do you want to tell us more about this constructive optimism philosophy?

Richard Socher: 00:56:05 Happy to. Yeah. I think optimists live longer. I think constructively optimistic societies progress faster and improve the state for everyone. I think you need to have a certain amount of optimism, but you cannot just be too



naive about the applications either, especially with AI, which is an omnius technology. We do need to carefully think about it and then construct that world. I think a lot of us get stuck in thinking that the future is kind of this prediction game like, oh, what's going to happen? But really it's a construction game. What future do we want to construct? And to realize that you can just do things. All the famous people in the world, they also just decided to do something and then went ahead and did it. And so I think realizing that agency is going to be more and more important in this world.

00:57:01 I think AI could also be a power to disenfranchise people and so on, but it can also be an incredible power for entrepreneurship because if you can all of a sudden delegate all of these little tasks that you have to do as an entrepreneur to an AI, you can just live out your ideas faster and try out different business ideas and so on. And so I think ultimately constructive optimism is I think a new, almost a philosophy, maybe it's a political philosophy, but it's also life philosophy to go out and assume you can try and build things, assume that you can change the world and then go ahead and try to build that change.

Jon Krohn: 00:57:45 Love it. Love that message. Let's move on to audience questions now. So as with many of the well-known guests that I have on the show, I post on LinkedIn that this guest will be coming up. And so got a couple of great questions here. One is from Jepson Taylor, who is, he's been a guest on the show more than anyone else, pretty well known in some data and AI circles for some years now. And I feel like the answer to this question, you could probably just even give a simple yes based on everything you've said in the episode so far. But Jepsen asks, the way that he asks it is kind of interesting, even if the answer is a simple yes. And so he asks, "Will we reach zero shot innovation?" Which means will models become so powerful that they can produce the missing knowledge



with no agent iteration, but a single inference with something like a flag of search set equal to one, even if that knowledge has never existed in the training set?

00:58:44 And in addition to knowing whether you think it's possible, which I feel pretty confident you think is possible, what's the timeline like on that?

Richard Socher: 00:58:51 It's actually a really deep question. I'll try to answer it quickly, even though it's hard to answer quickly. I think a lot of people assume you just have this eureka moment and comes out of nowhere. The truth is most innovations combine some smaller innovations from the past. And if you think about the combinatorial space of all possible ideas of math and physics and chemistry and biology and neuroscience and AI and algorithms and complexity theories, you combine all these, there's so many combinations that many combinations of these have been completely in areas of these kinds of combinations have been completely under explored and unexplored. And so you may say, oh, this idea like Einstein's relativity theory is one of those major theories, but it also sat on top of calculus and like Newton. And there are a few basic building blocks that would have been made it impossible without them.

00:59:49 And then other people finding close form solutions for some of Einstein's equations. And so I think there is this vast space that under explored and many of those innovations will seem fairly novel. We know the basic building blocks of amino acids, but to create completely new proteins can still have an incredible impact on humanity. So there's that. And then there's the question like, will we get to a place where it creates completely new paradigms that shift an entire way of thinking about something? I think that will also be possible and it'd be more possible if we first let it build these very complex systems with large neural networks. And then we can, when they get really good at something, then we can try



to interrogate them and understand and dissect how this neural network actually came to a conclusion. And we've been able to do that for very large neural nets where we see, oh, this neuron usually fires when we do this and that, and you can project it down to fewer dimensions to visualize it better.

01:00:51 And so you're right, the answer short is yes, but I think people underestimate how much is possible by just recombining all possible ideas that humanity has come up with in the past.

Jon Krohn: 01:01:05 Great answer. Thank you. I'm glad that there was more to dig into there than just a binary yes or no. Final audience question for you here. So I have an adjunct faculty role at Tulane University in New Orleans where I teach courses on AI and Professor Jon Renny, he's the main instructor on a course that we're doing together this term. So he had a question for you. He says that him and a colleague were able to, in one hour with Claude Code, do the amount of work that until recently would've taken a team of urban planners in his case, working with computer scientists, probably a couple of years and a half million dollar grant, and now it's done by CloudCode in an hour. And so his question is kind of related to how this will impact research in universities because why would a faculty member want to be mentoring younger students, PhD students, postdocs, when they could be just sitting with their ClaudeCode terminal?

Richard Socher: 01:02:06 I had a similar epiphany recently when I just worked... I used to do research over the last few years, mostly through other people. I asked, I gave them ideas, I discussed the ideas with them and they would go and implement them. And now I can just discuss those ideas directly with AI and it goes, implements them and balance it. And so over one weekend, I can now run 10,000 experiments on ideas that I have with no other human involved. And so it is mind-blowing how much more



productive we can be. I think one, like gyms of the mind, universities will have to create new minds as well, and that is part of their job and that job will change, but I think not in a way where we don't do it at all, number one. Number two, we will all be able with enough agency act on these higher levels of abstraction and become more and more like National Science Foundation directors, program directors who say, "I want to see this change in the world." Politics and people are asking us to invent these kinds of things and then everyone will collaborate with their AIs.

01:03:13 And so I do think it's important for researchers to start using these tools and to teach how to use these tools well and then realize that, wow, having really good ideas and the right level of management skills to bring those ideas into the world together with agents, that will be a really important skill that we have to teach everyone at university.

Jon Krohn: 01:03:39 Still some things for humans to step in and do for a few years to come.

Richard Socher: 01:03:43 For many years to come. I think to have truly unique and interesting ideas is really important. And then maybe I'll also end on another positive note, which is there will always be certain kinds of software that will stay and certain kinds of jobs that will stay. In terms of software, you'll have software that requires and has unique data, will continue to be relevant, that has network effects and marketplaces and so on will be relevant. You can vibe code. You can vibe code the Instagram app, it's just no one is on your little app. And so you don't have that network effect in it. And then brands, brands will be important. And in terms of jobs, it will always be like entertainment, health and research and pushing the frontier of knowledge forward. And then of course lots of jobs in and around AI.



- Jon Krohn: 01:04:36 Very cool. Thank you for this amazing interview, Richard. Before I let you go, how should people follow you for your thoughts after this episode? What are the best ways to follow you on social media or to get more information about Recursive or buy your book?
- Richard Socher: 01:04:52 Yeah, buy the book on Amazon Eureka Machine. I'm on X Twitter @Richard Socher.
- Jon Krohn: 01:04:59 Fantastic. Thank you so much for taking the time out of your very valuable day. Appreciate
- Richard Socher: 01:05:04 It. Thanks for having me. Thanks for listening everyone. Bye.
- Jon Krohn: 01:05:07 Fascinating episode today. In it, Richard Socher detailed how AI keeps replacing handcrafted human expertise with scale, first with features and then with architectures and how his company Recursive aims to automate the scientific method itself and point it at AI. He talked about the four pillars of his Eureka machine, existing world knowledge, scientific measurements, simulations, and real world experiments, all orchestrated by a swarm of agents. He talked about why hallucinations can be a feature for science since they let models explore a vast space of ideas while factual questions still demand grounded sighted research. He talked about how AI could turn biology, neuroscience, and even economics from sciences into engineering disciplines, why even if RSI arrives within a year or two, the need for compute and experimentation rules out an overnight takeoff and why software with unique data, network effects or a strong brand will endure along with jobs in entertainment, health and research.
- 01:06:02 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 Richard's



social media profiles as well as my own at superdatascience.com/1037.

- 01:06:16 That's today's episode number number 1037. And if you'd like to engage with me in person as opposed to just through social media, I'd love to meet you in real life next week at the Open Data Science Conference, AI West, O-D-S-E AI West, which is running from October 27th to 29th in San Francisco. I'll be delivering a keynote on how to thrive in the agentic AI era, amongst other fun duties like hosting fireside chats and emceeing. I'd love to meet you there. All right, thanks to everyone on the Super Data Science Podcast team, our podcast manager, Sonja Brajovic, media editor, Mario Pombo, our partnerships team Natalie Ziajski, our researcher, Serg Masís 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.
- 01:07:08 You can support the show by checking out our sponsor's links, which are in the show notes. The clicks really do make a difference for us. So think about giving those links a click. And if you yourself are ever interested in sponsoring an episode, you can get the details on how at jonkrohn.com/podcast. Otherwise, share this episode with folks who would love or be interested in the forward-looking views of Richard Socher. Review the show on your favorite podcasting app or on YouTube. Subscribe if you're not already 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. Until next time, keep on rocking it out there and I'm looking forward to enjoying another round of the Super Data Science Podcast with you very soon.