

**SDS PODCAST  
EPISODE 697:  
THE (SHORT) PATH  
TO ARTIFICIAL  
GENERAL  
INTELLIGENCE,  
WITH DR. BEN  
GOERTZEL**



- Jon Krohn: 00:00:00 This is episode number 697 with Dr. Ben Goertzel, CEO of SingularityNET. Today's episode is brought to you by AWS Cloud Computing Services, by the AWS Insiders podcast, and by Modelbit for deploying models in seconds.
- 00:00:20 Welcome to the SuperDataScience podcast, the most listened-to podcast in the data science industry. Each week, we bring you inspiring people and ideas to help you build a successful career in data science. I'm your host, Jon Krohn. Thanks for joining me today. And now let's make the complex simple.
- 00:00:51 Welcome back to the SuperDataScience podcast. Get ready for a mind-blowing conversation today with the visionary Ben Goertzel. Dr. Gurtz is CEO of SingularityNET, a decentralized open market for AI algorithms that aims to bring about artificial general intelligence, AGI, and therefore the singularity that would transform society beyond all recognition. Dr. Kurzel has been chairman of the AGI Society for 14 years. He's been chairman of the foundation behind OpenCog, an open-source AGI framework for 16 years. He was previously chief scientist at Hansen Robotics, the company behind Sophia, the world's most recognizable humanoid robot. He holds a PhD in mathematics from Temple University and held tenure-track professorships prior to transitioning to industry. Today's episode has parts that are relatively technical, but much of the episode will appeal to anyone who wants to understand how AGI, a machine that has all of the cognitive capabilities of a human could be brought about and the world-changing in impact that would have.
- 00:01:50 In this episode, Ben, details the specific approaches that could be integrated with deep learning to realize in his view, AGI in his view, as three to seven years. He talks about why the development of AGI would near-instantly

Show Notes: <http://www.superdatascience.com/697>



trigger the development of Artificial Super Intelligence, a machine with intellectual capabilities far beyond humans'. He details why despite triggering the singularity beyond which we cannot make confident predictions about the future, he's nevertheless optimistic that AGI will be a positive development for humankind. He talks about the connections between self-awareness, consciousness, and the ASI of the future. And with a admittedly wide error bars, he speculates on what a society that includes ASI may look like. All right, you ready for this astounding episode? Let's go.

00:02:39 Ben, welcome to the SuperDataScience podcast. It's awesome to have you on. Where are you calling in from today?

Ben Goertzel: 00:02:46 I'm on Ashland Island off the coast of Seattle.

Jon Krohn: 00:02:51 Oh, nice. Well, thank you very much for coming on the program. We actually hadn't met before but you're well known entity in the artificial intelligence space, and so I've had my eye on you for a while and we were delighted that we were able to get you on. I know that this is going to be a fascinating episode. So, you're creating a benevolent decentralized AGI at SingularityNET. Could you parse those terms for our listeners? So, AGI probably most listeners know, but even a quick intro to that just in case could be good. But then these ideas of a decentralized AGI and a benevolent AGI, I think those terms we should definitely dig into.

Ben Goertzel: 00:03:39 Absolutely. And I want to add none of these terms has an absolutely definite, well-defined meaning. These are things that we're fleshing out as we, as we go along, which I think is a, is a feature, not, not a bug. I mean, just as the definition of life is something biology is flushing out as it moves forward, which is just fine and doesn't stop people from doing synthetic biology or

molecular biology or whatnot. So, in terms of AGI to start, that's a term that I launched onto the world in 2005 with a book titled Artificial General Intelligence, which was a edited book of papers by different researchers doing research aimed at AI that can generalize in the sense of making big leaps beyond its programming. And, and its, and its training. And there's a whole mathematical literature aimed at defining exactly what does AGI mean.

00:04:41 When you start to formalize it, you find, okay, humans are not totally general intelligences. Like, we can't run a maze in 700 and 1006 dimensions very well, right? We can, we can barely figure out what each other are thinking and feeling most of the time. On the other hand, we can leap beyond our training and programming much better than a worm, a dog, or ChatGPT for that matter, which, while powerful as powerful mostly because it's training data is so big, it doesn't leap that far beyond its training data. Decentralized is a word that is big in the, in the blockchain and crypto sphere. And what it really means is not having any central owner or controller who's sort of the puppet master and pulling all the strings of the system. So, distributed computing is a more limited notion. I mean, Google or, or Microsoft within their server farms uses large scale distributed computing, but, there's one corporate entity controlling it all.

00:05:51 And on the computer science level, there's often a central controller sort of disseminating stuff among all those different machines and controlling them. Decentralized control generally would need a distributed infrastructure, but it goes, goes beyond that, right? And then blockchain is one sort of tool that can be used to enable decentralized control of a network of compute processes. But of course, you could talk about decentralized control of groups of people as well, as well, right? And so that's gets into anarchist political theory and a whole bunch of other fun topics. But I mean, clearly, I mean, US is a less

centralized system than mainland China or than an ant colony, right?

00:06:43 So, that, that, that notion is, is widespread in any sort of multi-agent system. Benevolent, of course, there's a whole field of ethical philosophy and people struggle to define these things in a precise way. But what we're looking at here is, in an AI context, is making an AI that is broadly speaking for the good of sentient beings rather than purely or primarily for the good of the small number of parties who owned and created it. And of course, not many entities are out there saying, we want to make evil AI that will kill and, and torture everybody. But you do have, I mean, a country creating AI to increase its own position relative to other countries where you have a company creating AI with the proximal of maximizing shareholder value, that's not necessarily malevolent. I mean, companies and countries seeking their own interest have done great things in the world.

00:07:55 On the other hand, it's different than creating AI with a goal of broader good behind it. And this often slips in the real world, right? Like OpenAI, who's everyone's favorite target to beat up on now, and I mean, they, I mean, they've done, they've done incredible things, obviously. So, I mean, my hat's off to them for, for having the balls to launch, you know, the first really powerful generative language model upon the world. On the other hand, they started off with a rhetoric of being open-source and nonprofit for the good of the world. Now they're closed-source and closely allied with a mega-corporation who's closely allied with the intelligence organizations of a particular country, right? So, that, that, not that they're bad guys or want to hurt people, they're good people who want the good of the world, but this illustrates how just the realities of human society and economy that you sort of shift kind of, it's a slippery slope from you start out wanting to save the world and you end up, well, the

easiest way to build stuff is to serve a sort of more narrow set of goals, right?

00:09:09 And that, that's the challenge with the benevolent part of, of benevolent decentralized AGI. Not, not so much bad guys who are like, I want to build killer robots to exterminate everyone, but just the tendency to shift toward it a narrower perspective to get done.

Jon Krohn: 00:09:30 Yeah, one of my favorite cutting quips about OpenAI is to call them ClosedAI.

Ben Goertzel: 00:09:38 Yeah, you have, in mathematics you have the notion of a CLO and set, right? So, that's they're a bit like that. But yeah with GPT-4, really, they didn't really tell you what's happening, but behind the scenes, right? I mean, you can, you can guess what sorts of things they probably integrated with the base transformer model, but they, they really don't, their paper doesn't actually tell you what's going on. I mean, even short of open-sourcing the code, they have, they haven't really disclosed what the algorithmics are, which is pretty far in the opposite direction, right?

Jon Krohn: 00:10:54 Are you stuck between optimizing latency and lowering your inference costs as you build your generative AI applications? Find out why more ML developers are moving toward AWS Trainium and Inferentia to build and serve their Large Language Models. You can save up to 50% on training costs with AWS Trainium chips and up to 40% on inference costs with AWS Inferentia chips. Trainium and Inferentia will help you achieve higher performance, lower costs, and be more sustainable. Check out the links in the show notes to learn more. All right, now back to our show.

00:10:58 Yeah. So, with SingularityNET, the hope is that your SingularityNET ecosystem will be able to maintain more

benevolence, more openness, I suppose, than these other kinds of entities like AI.

- Ben Goertzel: 00:11:15 Well, yeah, that's right. And I think open-source is doing great in the AI world, generally speaking. I mean, the vibe in AI is most algorithms do get open-source. Most innovation happens in published papers that are published on Archive or elsewhere. And even though GPT-4 is the smartest LLM out there at the moment, I mean, others are rapidly on its heels with, with open open-source models. And the VC community, which I have a lot of issues with, has to their credit, opened their minds to open-source business models and is putting a bunch of money behind people building, building, open, open-source AI tools, right? So, I mean, I mean, I think there's, there's a lot of hope in that regard, but I do think open-sourcing the code isn't all you need to do, right?
- 00:12:19 So, there's code, there's the data that code is, is, is trained on, and where that comes from, what's the ownership model, what's the compensation model behind it? Then there's the sort of live network of machines, which is running the AI code and who's hosting and owning and controlling those machines. And all these things need to be done in a democratic and decentralized way. But I think the success of open-source code and open science in the AI field now gives a great start in that direction. I mean, along with the software tools from the SingularityNET ecosystem and others that handle some of these other aspects of the problem.
- Jon Krohn: 00:13:09 Yeah. Fill us in more on the SingularityNET ecosystem specifically and how it solves some of these problems.
- Ben Goertzel: 00:13:15 Yeah, so SingularityNET, we started in 2017 with the goal of making a decentralized platform for AI. And both the specific technology and the nuance of the goal have shifted a bit since then just because both the blockchain



and AI worlds have shifted a bunch since that, since that time, right? So, the, the original notion in 2017 was, okay, let's create a blockchain based platform so that anyone who creates an AI can put an AI online and then they can put it on their machine, they can put it online anywhere. They connect it to the SingularityNET network of other AIs out there. And all these AIs can outsource work to each other. They can talk to each other, they can collaborate, solving customer problems. And the intelligence of the whole can be greater than the intelligence of the, of the parts, right?

00:14:17 And we, we rolled this out as a platform, a sort of decentralized multi-agent system for, for AI on the Ethereum blockchain initially, because that was sort of the only one there that supported smart contracts, which are, you know, they're neither smart nor contracts really, but they're persistent scripts that allow secure, decentralized control of, of distributed software processes. And so we, we built that platform, I would say it was a success technically in as much as it went. It didn't, to be honest, get as much traction as we'd like, that we got a bit of traction. And I think blockchain technology didn't mature as fast as we thought it would. So, I mean, you just had super high Ethereum gas costs and slow networks and so on. So, from an AI developer's point of view, it's like, why would I put my AI on this really slow infrastructure, you know, so people could pay with crypto, but not everyone does, and for philosophical reasons, but that doesn't always override the practical irritation.

00:15:33 So, I mean, people, people using AI for decentralized finance or something like the platform, because they have crypto they want to pay with and they're already bought into the crypto ecosystem, right? But we didn't, we didn't get as much penetration as we hoped in the broader AI sphere for, for pretty clear reasons. So, we pivoted a little





bit in 2020, 21 toward creating our own projects, addressing AI needs in various vertical markets. Leveraging our decentralized platform and at the same time plunging into building out the whole decentralized ecosystem beyond just SingularityNET platform, right? So, we, we created, for example, a project called Rejuve that uses the decentralized AI on the platform to analyze, you know, bio signals data and medical data that people upload and uses AI running decentralized on singular net platform to tell, tell you stuff about your own, your own body and your progress toward, toward, toward longevity.

00:16:43 And we, we created SingularityDAO, which uses AI running a SingularityNET platform to do crypto trading, various sorts of decentralized finance, and a few other projects along those lines. Something called Mindplex, that uses this decentralized AI to help with that with decentralized media. So, it's more, okay, well let's, let's ourselves build stuff that shows what a decentralized infrastructure can do for AI, because we're doing it because we, we grasped the long, long-term vision. And then, you know, then in the last six months we have this whole LLM revolution, which is fascinating and interesting, and we can dig more into that a a little later. I mean, I think LLMs are a breakthrough. I think they're not yet getting us to AGI. I think we can enhance them a lot using other AI things that my team knows about, like neurosymbolic AI and evolutionary AI plugged into LLMs.

00:17:52 But the amplification of LLMs for the decentralization of AI and for platforms like SingularityNET are significant, right? Because it means that it seems like a lot of the progress in AI is going to be little apps that leverage large language models and the successors for large language models to do things rather than small standalone AI agents. So, what that means is, okay, if you want to make AI decentralized, a. you have to make LLMs and their descendants be the neurosymbolic LLMs, whatever, you

have to make these decentralized. And then what your whole agent system is doing, it's like an App or dApp ecosystem of little AI agents that they may interact with each other in a sort of heterogeneous way, but they're also making a lot of API calls into these decentralized LLMs or LLM successors, right? And that, that's something we've been thinking through quite a lot and trying to build infrastructure for, to compliment the core SingularityNET blockchain based multi-agent system platform.

00:19:10 So, we, we made a platform called NuNet, which lets you sort of contribute processing power that you have to the decentralized AI network. And then in an LLM context, you know, you really need a powerful server farm to train large language models. But if you're doing fine tuning or you're doing prompt tuning or you're doing learning of this symbolic portion of a neural-symbolic large language model system, you can do that on your phone or laptop or something, perhaps in a way that's centered on your own data, which is on that device. And we can use NuNet to help make that aspect decentralized. Then, then we've launched our own layer 1 blockchain project called Hypercycle, which is a ledgerless blockchain. So, it goes beyond the decentralized ledger aspect of Ethereum. Another commonplace for blockchains, and again, that's, that's aimed at making it not be a slow and expensive thing to put the blockchain underneath your AI application, be it LLM oriented or, or otherwise.

00:20:20 And so each of these things, NuNet and Hypercycle are their own projects, which we've separately capitalized and sort of spun out of the SingularityNET Foundation, which was the initial entity that that launched SingularityNET platform. And then for the large language model aspect, we've spun off a company called Zarqa, whose goal is just to build, build some large language model supercomputers, use them to train large language

models, but on as much of a decentralized infrastructure as we can, right? So, I mean, we'll use a supercomputer when we have to, but then for fine tuning and prompt tuning and symbolic piece, we can use NuNet and SingularityNET to fully decentralize those aspects.

00:21:10 But also the core training of an LLM doesn't have to be just on one server farm that one entity owns, right? It could be, it could be split across, you know, Zarqa's own server farm, but then a bunch of, say former Bitcoin mining farms that want to put a bunch of their machines into that. And then, then we've spun off a company called TrueAGI, which is oriented toward OpenCog Hyperon, which is a cross-paradigm, like neural-symbolic evolutionary AI paradigm bringing LLMs together with other stuff, which can then run on this decentralized AI platform. So, I mean, I think what we're trying to, trying to think on our feet to make decentralized AI and AGI actually work in this rapidly evolving AI and blockchain landscape, right? And I've sort of come to the conclusion, I've come to the conclusion to make decentralized AGI really work, I mean, pretty much we have to launch something that's way smarter than ChatGPT and launch that on a decentralized infrastructure, right?

00:22:22 And if, if you do that, everyone will use it because it's smarter. And lo and behold, it also happens to run on this decentralized infrastructure and no one, no one will b\*\*\* about that fact, even if it's not their main, their main purpose for, for using it, right? I mean, just like people are happy to use stable diffusion, they like the fact that it's open-source. On the other hand, if there was something proprietary that worked twice as well, they'd probably use that. Even though it was closed-source, right? So, that, that's sort of what my focus is now. Like how do we use LLM symbolic reasoning, evolutionary AI, all these different things together to make something way smarter than ChatGPT, and then we can roll it out on our

decentralized infrastructure and then curate an app ecosystem around that, doing all sorts of things serving different, different vertical markets.

- Jon Krohn: 00:23:19 This episode is supported by the AWS Insiders podcast: a fast-paced, entertaining and insightful look behind the scenes of cloud computing, particularly Amazon Web Services. I checked out the AWS Insiders show myself, and enjoyed the animated interactions between seasoned AWS expert Rahul (he's managed over 45,000 AWS instances in his career) and his counterpart Hilary, a charismatic journalist-turned-entrepreneur. Their episodes highlight the stories of challenges, breakthroughs, and cloud computing's vast potential that are shared by their remarkable guests, resulting in both a captivating and informative experience. To check them out yourself, Search for AWS Insiders in your podcast player. We'll also include a link in the show notes. My thanks to AWS Insiders for their support.
- 00:24:10 Yeah. So, let's dig into some of those technical terms like neurosymbolic systems, evolutionary learning knowledge graphs in one second. But just before we get there, I want to kind of recapitulate back to you what I took away from everything that you've covered so far. So, the idea with SingularityNET...
- Ben Goertzel: 00:24:27 Makes sense. I went on a long time. Yeah. So, yeah.
- Jon Krohn: 00:24:30 No, it was great. So, with SingularityNET initially, and then these other spinoffs out of the SingularityNET Foundation, like NuNet, Hypercycle Zarqa, TrueAGI, the idea, of course, is to have benevolent, decentralized AGI like you defined at the outset of this episode. And it's interesting how this journey started with leveraging existing blockchain technologies like Ethereum, but then along the journey, realizing that those underlying technologies, blockchain technologies weren't evolving

quickly enough to be able to get the kind of economics that you'd require to be handling the very large data sets, the very large models associated with modern AI. So, it sounds like you went off to, you broadened to be able to start building this infrastructure yourself. And in so doing, you've also created infrastructure that can, that can support other kinds of use cases, not just AI, but also things like financial transactions, like longevity. And so, yeah. So, that sounds like, you know, a really exciting project and lots more to come. In terms of your vision for how we can realize AGI you talked about combining neural networks with other kinds of approaches like neurosymbolic systems, evolutionary learning, knowledge graphs. Could you dig into those three approaches and describe how they can...

- Ben Goertzel:      00:26:12      Absolutely. So, yeah, I think you know, as most of the people listening to this podcast probably realize, you know, the AI field has been around since the middle of the last century. I mean, the name was invented in, I guess the late fifties but goes back a couple of decades before that. Probably Norbert Wiener's book on Cybernetics was the first place it really laid out to have AI as a discipline for making machines do the kind of thinking that brains can do. That would've been in the mid-forties sometime. Like I'm old, but not, not that old. I was born in the late, in the late sixties, right? And during that long history of, of AI field, a lot of different AI approaches have been put out there. I mean, neural nets have been there since, since the forties, but logic systems have been there since 1960s I would suppose, and evolutionary learning that tries to emulate the process of natural selection to do thinking. This has been there since the early, early to mid-seventies.
- 00:27:21      Hosts of different AI approaches have been around for a while and deployed in commercial systems across various vertical markets with sometimes great success,

sometimes less. So, right now, one of the interesting things we see with the whole LLM revolution is just taking stuff. It's not that different than what was done before. I mean, decades ago, even with some minor tweaks, deploying it at way larger scale makes some quasi-magic happen, right, and it's pretty cool. I mean, I remember in the late nineties, mid-nineties, I was teaching cognitive science in university before I bailed on academia to go into industry. I was teaching neural networks with a neural net with like 35 neurons trained using recurrent backpropagation on a decent, like Sun Unix workstation, take three or four hours to train that network with 35 neurons, right?

00:28:24 And I would tell the students, well, okay, this is 35 neurons. Your brain has like a hundred billion neurons, but it's got massively parallel infrastructure. Like as hardware gets better and better, we're gonna be able to train these software networks, you know, similar to how, how the brain does things. At the time, I was working on a connection machine from Thinking Machines Incorporated, which had like 128,000 processors, like MIMD parallel, 128,000 autonomous processors doing different things that could train a neural net much faster, right? They've stopped making that kind of hardware. We got, we got GPUs instead. But in the end, you know, what I told those students is basically accurate, right? I'm, I mean now I, now we can trade, train a s\*\*\*load of neurons in a neural net. It's fast and it does amazing stuff, which is just what I thought would happen.

00:29:23 Honestly, I thought that was gonna take five or 10 years, not as, not as many years as it has since the mid-nineties, but. And there's been architectural changes, right? So, transformer neural net, if you go back to the Attention Is All You Need paper. A lot of what happened there is you took recurrent neural nets, you stroked out some of the recurrence and replaced it with an attention



mechanism. Now that, that decreases the computational capability of the network. I mean, in terms of just theoretical computer science, a vanilla transform, a neural net doesn't compute all the kinds of things. It can't by-simulate all the kinds of automata that a recurrent neural net can. I mean, if you give it an external memory and a scratch pad, it can, it can kind of do it, but it not natively, right?

00:30:08 So, you're, you're weakening the computational power of recurrent neural nets by replacing recurrent layers with attention, but you're making it easier to train at a huge scale, right? And that's, that's really interesting. We don't yet know exactly how much of the loss that you get from getting rid of recurrence can be gotten back by other, other tricks within the scope of transformers. But it's clear, like we can do so much amazing stuff by sort of scaling up a fancy version of the multilayer perceptrons that we're working with in the late, in the late sixties, right? So, now you go to other AI paradigms like logical theorem proving, right? Well, in some domains we already see what can be done by scaling them up in modern computing infrastructure. So, you know, my oldest son Zarathustra just submitted his PhD thesis in Technical University of Prague on using machine learning to guide automated theorem proving.

00:31:16 So, I mean, if you take a corpus like Mizar, which is a list of huge number of mathematical theorems going to, going to the PhD level and beyond, pretty rapidly a machine learning guided theorem prover can prove like 80% of them just zing, zing, zing, zing, zing, right? So, I mean, certainly more than I can do in a few hour's time, and my PhD she was in math, right? Now, they're not good at making up amazing new math theorems, which is what mathematicians have, have fun doing. But they can prove things very quickly and very well. And again, the core concepts behind today's automated theorem provers,



they're not that different than the core concepts behind automated theorem provers from 20 years ago. It's just, you know, we've optimized the code behind these, but we also have just much faster computers to run them on.

00:32:10 And having these, the internet where you can make a corpus of all math theorems and just try stuff over and over again on this corpus, tune the parameters of your theorem prover, and the machine learning, that that, that, that guides it. I mean, this, this has allowed conceptually similar ideas to the ones from decades ago to be refined in such a way they're super, super good at theorem proving, right? And I think the same holds in a, in a bunch of different, of different areas. So, we could look evolutionary learning, the idea there is assimilating the process of evolution by natural selection like mutation, and mutation and recombination and selection of the fittest entities. So, the basic dynamics of evolution of, of populations, of biological organisms, you're simulating inside the computer. One example of evolutionary algorithms stopping short of AGI, which I'll talk about in a moment, but just to understand the impact of the modern ecosystem on making evolutionary algorithms work better than they, than they used to.

00:33:20 I mean, in the mid-nineties, I used genetic algorithms, floating-point genetic algorithms to evolve the coefficients of iterated function system fractal generators to generate sequences of melodies, musical notes. And I then used a rule-based AI system to modulate the timing. And I was able to generate some quite, quite cool music. I'm an either modern classical or I made it generate like long rhythm guitar and league guitar stuff that sounded like, you know, epic versions of the middle of Master of Puppets by Metallica or something, right? It was, it was, it was cool, but I mean, it was, I couldn't get the timing to be interesting from the, from the evolutionary algorithms. So, I made a rule-based system based on some music

theory to, to do the, do the timing rules. And I mean the tomber, the tomber I just did in a standard computer music way, I used the GA just to evolve the practical, efficient to generate a series of notes, like, like a, a MIDI file for those who do computer music.

00:34:27 So, that, that was interesting. And the mutation and crossover of a genetic algorithm, it lent some creativity to the proceedings, right? So, instead of, at the time when people were doing an AI for music was like Markov modeling, like you just generate a probabilistic model of this series of notes and generate new series of notes by instance generation from that distribution. And I mean, that's, that's cool, but it gives you something that sounds like a second-rate version of the stuff in the training corpus, right? And genetic algorithms generated a bunch of garbage, but they also generated stuff that was new sometimes. So, now, now I've been revisiting this area, but in a modern way, right? So, you can take something like MusicGen, which Facebook has conveniently released recently into the open-source, which is great, unlike Google that did not release their music LM into the open-source.

00:35:24 So, good for you Facebook, bad for you Google in this particular case, right? And so music LM, you can give a text prompt and optionally an audio file, a music file, and it generates new music conditioned on the, on the music file, guided by the text prompt, right? And that's, it's very nice. It generates stuff that sounds like music. Now what I found is just using its prompts, it generates music, but it's sort of cliché music that doesn't interest me musically, even though it's very competent. If I use my own weird music as the melody prompt, then, then it generates stuff that I find interesting guided by the text prompt. Now, since it's open-source, if you plunge in and see how they use the melody prompt, it's very crude and simplistic. So, I'm currently mucking around in there,

working on some better ways to do the conditioning on the melody prompt. If I get anything, awesome, we'll release that open-source also.

00:36:26 But what I'm playing with with genetic algorithms is use a GA to evolve prompts, right? So, using a genetic algorithm to do mutation and crossover, and then probabilistic instance generation on prompt space. So, you have a genetic algorithm, you're mutating prompts, you're taking a prompt and combining it with another prompt. Then as a human, you listen to the music that came out from a given prompt. You rate it for how much you like it, and then you feed your evaluation back in as a fitness evaluation to the genetic algorithm for prompt evolution. Then, then you can code rating functions which embody like a mathematical model of musical aesthetics. And you use that, use that for fitness estimations.

00:37:11 So, if you're, if your rating function says this sucks, you don't listen to it. If your rating function says it's good, then you listen to it and give it a rating as a human to feed back into the, as a fitness function of the genetic algorithm. So, the thing is, using this spiritually, it's like similar to what I was doing in the nineties, right? Using a GA to try to evolve some parameters that guide another process that generates music. You can listen to it and rate it and try to get something creative, but now it sounds like great diversity of music in any genre with any number of instruments, right? And it's, that's super interesting. I mean, this is because you have a model like music LM, which is bottoms out on all sorts of like sequential models of music and LLMs doing the text and music mapping and so forth. So, we've got a lot of underlying technology that lets the same concept of evolutionary learning of the coefficients of a music generation process work now just much more flexibly than in the nineties.

00:38:18 And now, like all the parts are done within the same AI process. Whereas in the nineties it was like, okay, use your GA and fractal generate melody, then find some other music theory way to generate timing, find some other way to generate tomber. Like now it's all in the same AI process. So, what we've seen many domains here is more data, more compute, same conceptual ideas, lets you tweak the specific algorithms to a way that makes things magic and make things, makes things super cool. Now, what I think is, you know, this is gonna let us actually make artificial general intelligence at the human level not too long from now, let's say three, three to seven years from now. And I think once you get to artificial general intelligence at the human level, if it's done in a non-stupid way, that's gonna relatively rapidly lead to artificial intelligence at the way beyond human level because...

Jon Krohn: 00:39:19 And Artificial Super Intelligence.

Ben Goertzel: 00:39:20 AGI at the human level ...Yeah, I mean, if the AGI is as smart as you or me and has full read and write access to its source code, I mean, it's gonna figure how to revise its source code and it's gonna uplift itself the Artificial Super Intelligence, which leads to a whole class of, of issues that we can get to if we have time. But let me, let me come back to logic systems and evolutionary AI and knowledge graphs and how I think they can potentially be used to make the leap from really cool narrow AI systems like ChatGPT and like the other things I've been doing to artificial general intelligence. So, first of all, I'd say nobody solidly knows how to make the leap from where we are now to human-level artificial general intelligence. I mean, it's a research question.

00:40:13 Last week in Stockholm, we had the four-day-long AGI23 conference, which is an artificial general intelligence research conference. And I've organized that conference

every year since like 2006. This was obviously more LLM-heavy than most, but you had a bunch of other sorts of AGI ideas and a bunch of hybrid LLM logic and other sort of system talks. But there's, there's a broad spectrum of ideas about how to make the leap from here to AGI. And I'd encourage people to look at the, on, on YouTube, the video proceedings in that conference. I mean, the first day we had a workshop on OpenCog Hyperon, which is my own primary AGI attempt now, but the main body of the conference had talks by all sorts of other people. And if, if you're interested in large language models, if, if you look at Noah Goodman's talk showing how good large language models are doing in elementary math, and then [inaudible 00:41:17] talk discussing how terrible large language models are at doing slightly more advanced math and logical reasoning. It, it's an interesting sort of counterpoint.

00:41:28 So, one approach to making AGI will be to, you know, try to make GPT-7 or something, right? I mean take, take a large language model, plug more stuff into it, right? Plug Wolfram | Alpha into it for calculation thinking, right? I mean plug, plug some sort of long-term memory into it and some sort of better working memory into it. So, it keeps the thread during a whole conversation and has some coherent lifelong experience, right? Plug, plug the best, you know, video processing models you have in it so that it can intersect video and language fluently, which is almost but not quite there, right? So, for music, like right now, GPT-4 has never listened to music. No one knows the text associated to music, but I mean, you can make it, intersect it, co-train it with a different neuron net that's seeing in inside music. So, that would that, that, that would, that would be one approach. Now, no, I'm working centrally on a different approach, which is within a code base called OpenCog Hyperon, which is a new version of the OpenCog open-source AGI.



- Jon Krohn: 00:42:46 This was my very next question. It's perfect.
- Ben Goertzel: 00:42:49 Yeah, I mean the core idea there is you take this large distributed, potentially decentralized knowledge graph or more properly, it's a meta graph, not a graph, but same concept. I mean it means like a graph in a sense is nodes with links between them. A hypergraph is like a graph, we can have one link spanning five nodes or 20 nodes or something rather than links being binary. A meta graph is even more general because I have a link pointing to links or a link pointing to a whole sub-graph. So, we take, take this knowledge meta graph and we use it as a sort of knowledge meta representation where you can put all sorts of knowledge in the same meta graph. And that includes logical knowledge. It includes neural net distributed knowledge.
- 00:43:47 It can include programs. So, you can embed a program inside this knowledge meta graph. And of course, you know, inside a C++ compiler or a Haskell compiler, you have a graph representation of a programming language of a program, right? So, I mean, representing programs as graphs is not, is nothing new. But here it's the same graph used to represent declarative facts and beliefs, neural networks or which are themselves nodes and links, right, or programs. And then pieces of that graph can then be sort of brought alive by different software processes. So, if you have a program stored in the graph, an interpreter can grab that program and run it, and then the programs stored in the graph. What do they do? Well, they transform the graph, they pull knowledge out of the graph and they rewrite the graph, right? So, we made a special programming language called MeTTa, Meta Type Talk, M E T T A, which basically is a programming language encoded in the graph. And what it encodes is graph transformations, right? So, then basically you have this big distributed, decentralized self-modifying meta knowledge graph, right? And that's, that's a framework.



You could use that framework for a lot of different things. And what we're using it for now is doing probabilistic programming. We're using it to do probabilistic and fuzzy logical reasoning. We're using it to do evolution of programs that are stored as subgraphs and we're interfacing this knowledge graph with external software libraries running deep neural networks.

00:45:28 So, you can run a deep neural network totally inside the Hyperon meta graph, is just not as fast as running it in Torch or TensorFlow or something. So, what we're pretty much doing is, we have nodes in our knowledge graph without references within the torch compute graph. But then you, you actually do the neural net stuff in torch at the, at this moment just because it's, it's faster. I mean it may not always be faster, but it is right now. So, what you have here, you have the ability to do, you know, deep vision model or large language model stuff outside of the Hyperon system. You then have the Hyperon system that can drive and interact with and guide these neural systems. And you can then do logical reasoning in the Hyperon system and evolutionary learning in the Hyperon system. And all this is designed to be rolled out on the SingularityNET decentralized AI platform and leverage new net for compute resources, leverage the Zarqa decentralized LLM for language modeling, leverage TrueAGIs infrastructure for rolling out a APIs based on all this, and Hypercycle ledger-less blockchain, blah, blah, blah. Like, it's designed to leverage this whole decentralized tech stack that we've built in, in the same way that OpenAI is integrating this stuff with the whole Azure centralized tech stack, right?

00:47:02 So, what we're planning on doing with this over the next couple years, I mean, in the big picture it's everything, right, just because AI will eat everything. But there's, there's a few critical development directions we're looking at, I mean, the most commercial one is just trying to



make something that's twice as smart as ChatGPT by kind of deeply integrating a probabilistic logical reasoning engine with an open-source LLM. And we're, we're building our own open-source LLMs, but you could also do that with, with Llama or whatever other open-source LLM you, you want to work with

00:47:46 So, the core idea there is, make something that's smarter at multi-stage, logical reasoning that integrates linguistic with quantitative and relational data and make something that's more creative and sort of less banal in its creative productions by adding on a logic engine and evolutionary learning engine to an LLM in a framework that was designed for interaction of AI from multiple paradigms, right? So, that, that's one thing. I mean, if we can pull that off on the decentralized infrastructure, then, then we're, we're, we're doing very well in the number of dimensions, right? Now, another thing we're looking at, which is more of the research side, is controlling a bunch of little learning agents or like baby AGIs and not, not like the bogus thing that stole my term BabyAGI for a ChatGPT rapper, right, but I mean, making things that really are like young, young artificial general intelligence like AGI toddlers trying to learn about the world.

00:48:49 So, I want to make a bunch of little guys in a virtual world and we have our own virtual world we're building called SophiaVERSE, which will have a bunch of avatars that look like the Sophia robot rambling around. But I also want to have these little AGI toddlers rambling around in SophiaVERSE and just interacting, building stuff together, chatting with each other, learning as they go. And that I would like to also do this within little, like small toddler AGI robots. I'm not sure if we'll pull that off, but we're, we'll definitely do it in the virtual world and that, I mean, these may not be as smart in terms of passing exams and stuff as a ChatGPT initially, but the

idea is they're learning from the ground up, from experience. So, I think both of these are interesting.

00:49:35 I mean, ChatGPT has set the bar reasonably high in terms of practical functionalities. We want, we want to surmount that, surmount that bar by just plugging LLMs into our system. And I mean we can, we can now use LLMs themselves to translate natural language into higher order or predicate logic so we can feed our logical knowledge base with all the knowledge on the web and feed that into reasoning engines, right? So, that's cool. On the other hand, I think there's a certain element of creativity and an element of like being an autonomous agent and charting your own path for the world of understanding who you are and understanding who others are and what's the relation between yourself and others. There's a lot of cognitive things there that people do, which may be easier for an AI system to learn if it doesn't have like half-a\*\*\*\* versions of all the knowledge of the web and it's mind already, it may be easier for an AI to learn if it's really just making sense of itself and the world in a simpler setting like, like a young toddler does, right?

00:50:43 So, I think we can do both of those things with OpenCog Hyperon in parallel. One is more commercial, one is more pure research. I mean we're, we're also doing other stuff. We, the first thing we loaded into our OpenCog Hyperon distributed Atomspace space is a bunch of bio ontologies actually we've imported all knowledge about fruit fly genomics and a lot of ontologies of human genomics. And we're just, because I have some colleagues in the Rejuve Biotech project where they have all this DNA data from fruit flies that live five or eight times as long as normal fruit flies that we've evolved over 40 years to live a long time. And, they actually need an AI system like this to understand why the flies live so long. What that might mean for human genomics, right? So, we're, I mean

they're, they're using it hands-on. So, I mean that it really spreads out in every direction once you managed to integrate these different things together.

- Jon Krohn: 00:52:20 Deploying machine learning models into production doesn't need to require hours of engineering effort or complex home-grown solutions. In fact, data scientists may now not need engineering help at all! With Modelbit, you deploy ML models into production with one line of code. Simply call `modelbit.deploy()` in your notebook and Modelbit will deploy your model, with all its dependencies, to production in as little as 10 seconds. Models can then be called as a REST endpoint in your product, or from your warehouse as a SQL function. Very cool. Try it for free today at [modelbit.com](http://modelbit.com), that's M-O-D-E-L-B-I-T.com
- 00:52:23 Yeah, it's wild to me all of these different projects that you have and how they all work together, and how you've conceptualized them. There are so many technical things that I would love to have had time to dig into more things like the way that you're thinking about OpenCog to be able to have a meta graph that blends together differentiable learning versus more discreet declarative information. I think that that's a brilliant way to go towards realizing AGI as well as your SophiaVERSE of toddler AGIs. I think these are all it, yeah, it's all so fascinating.
- 00:52:58 But I also know that I have a limited time with you today and so I want to get into some big questions. So, let's assume that one of your approaches or some other AGI I approach works, you think it's three to seven years before we have AGI realized. So, what are the implications? And so, you know, you've written extensively and extensively about things like life extension, immortality consciousness, and the singularity event when AGI is realized ties into these kinds of things. So, you even just

talked about one example there where things like Drosophila fruit fly genomics can be studied through AI systems might help us understand how we can be extending our lives. But yeah, you've written a lot about broader things than that than just these specific applications. But in a world where we have you know, huge amounts of energy resources, access to artificial super intelligence systems yeah. What are the implications? And it sounds like it's gonna be happening in our lifetime.

Ben Goertzel: 00:54:11 Yeah, I think the implications of super intelligence are huge and hard to foresee. And it's, I mean, it's like asking, you know, nomads living in early human tribes, what civilization is going to be like. I mean, they could foresee a few aspects of it, but to some extent you just have to discover when you get there. And that's, this is probably an even bigger shift than that shift from tribal to civilized life, right? Because, I mean, once you have AGIs that are 10 times the smartest as people, I mean, science fiction has explored many of the potentials, right? I mean, you could upload your brain into a, into a computer. You get like wifi telepathy to bring us all into a, some decentralized Borg mind. You get like drones airdropping molecular nano assemblers, or femto assemblers in everyone's backyards.

00:55:03 They can 3D print whatever matter they want or whatever matter they can convince the assembler to build for them, right? I mean there's options just going way beyond our current culture and way of thinking. So, I mean, I think that's fascinating to think about. And the confidence bars just have to be drawn really, really wide. And I mean, I say the same about people worried about the risk of superhuman AGI. I mean, you can't squash that risk to zero in a rational sense, right, because there's just so many unknowns. There's also no reason rationally to assume the risk is super high either. I mean, we just

don't, we just don't know what's gonna happen. And you may find that exciting or scary depending on your personality type and whether you're optimistic or pessimistic about it really says more about your psychological or spiritual makeup than about anything else. I'm very optimistic guy, so I have a lot of fun in life, right?

00:56:21 I think, the period between now and getting a human-level AGI is easier to think about concretely and there's a lot of meat to grab onto there, right? We can, we can already see with ChatGPT, it's pretty easy to see that with fairly straightforward integrations and improve small improvements of this technology. A quite high percentage of what people get paid to do for a living can be automated, right? Including this interview for example, I could see how ChatGPT can't yet automate this interview, but could you do it with slight advances of the technology short of AGI? Probably like ask GPT-6 what questions to ask Ben Goertzel in an interview given what he's worked on and what's of interest to our listeners.

00:57:19 And then train the model on me and say, well, "what would Ben say to these questions?" You know, I do vary a little bit each time, but I have talked about most of these things before. I have a neural model trained on myself, which totally couldn't do this interview as, as well as me. Right? But, but yet given how fast these technologies are, are developing, it doesn't seem infeasible. Now if you take a topic I've never given an interview on before, but I'm interested in say integrating, integrating general relativity and quantum mechanics, right? So, I've published a few papers on physics, but not a lot. I would bet you need a full-on AGI to emulate me in an interview on particle physics. Cause I'm just gonna be coming up with a lot of stuff that I've never said to anyone but have thought about a lot. Right? But so that, there's a limit, but the limit may be pretty high-end, right? I mean this, what

we're talking about is quite high-end in terms of journalism.

- Jon Krohn: 00:58:25 Yeah. Even in that scenario, I think it would end up probably you wouldn't need an AGI it wouldn't necessarily be representative of your thoughts, but it, we could do that today by taking-
- Ben Goertzel: 00:58:39 [crosstalk 00:58:39]
- 00:58:39 It might take-
- Jon Krohn: 00:58:41 Or other people's thinking.
- Ben Goertzel: 00:58:41 Yeah, it would take my own, it would take my own ideas and integrate it with stuff other people thought. Then I would've to watch that podcast to see what my simulator had invented. Maybe it would change my own thinking, but it still wouldn't, it wouldn't be a substitute. And I think there's a certain level of individual creativity which you don't get from these sort of weighted averaging type systems. And I, I think that until you have a full-on human-level AGI, you're not gonna quite get to that level of creativity. Like, if you think about music as another example, I mean, I think we're not there yet, just like we're not there yet with journalism. I mean, right, right now, like Mindflex magazine and SingularityNET ecosystem, we've tried, you can't get LLMs to write an article as good as a person. I mean, it's, it's not, it's not there yet. You can ask the LLM to write a sketch of an article and then we'll look up a lot of stuff for you and save you a lifetime on research. But it writes boring articles. It doesn't, it doesn't have much into it, right? It's, it doesn't really write the stuff you'd want, you'd want to put out.
- 00:59:51 But, you can see how you might get there without having an AGI. And similarly in music. I mean, you can, you can

come up with auto-generated pop songs that sound like, you know, the B-side of a single or a random band you might hear in a bar. You can't come up with a really great song. You certainly aren't gonna invent an amazing new genre of music. Like if you took LLMs and trained them on all music up to 1900, it's never gonna invent like neoclassical metal or, or free jazz or something based only on data from music up to 1900.

01:00:30 So, clearly there's a level of inventiveness that we're not getting and we probably need AGI for, but this sort of thing is a small percentage of the economy right now. Right? So, when you think about like, what do you really need AGI for and can't do with LLMs plus narrow AI advanced and scaled up a bit, I mean, there's a couple categories I think of, I mean, there's jobs that are just about human connection, like a preschool teacher or a nanny or a psychotherapist. I mean, it's about humans helping humans to be more human. And that's, that's fundamentally human, right? I mean, you, so you're, I mean, seeing live music as an element too, sometimes you want to, you want to see another person play, right? You don't, you don't, you could already hear a recording. If you're gonna see a robot play, maybe you might as well listen to a recording after the novelty.

Jon Krohn: 01:01:26 Yeah, I think it was famously The Gorillaz. The Gorillaz, I think there were like riots when they performed, they wanted to perform their concert behind a screen and people hated that.

Ben Goertzel: 01:01:38 Yeah. Yeah. I mean, sure you, you want that human feeling. You want to see the guy screw up, you want to see him get excited and passionate and you emote, right? And that, that's, it's amazing. Then, then there's in-depth scientific and technical innovation, which almost by definition is about taking multiple difficult steps beyond the state of the art, which I think current LLM



architecture doesn't do. It pretty much recycles the state-of-the-art, right? I mean, LLMs can't even do high school math or like an undergraduate economics exam, let alone radical, innovative, innovative science. Right? So, then there's cutting-edge art, I mean, radical artistic innovation for the same reason. It's not just recycling stuff, right? Then strategic thinking, I mean, what it takes to be a really good startup CEO for that matter or, or to be really good at the part of your job where you're figuring out who do you want to interview, who might not, might not be mainstreaming on everyone's radar yet.

01:02:51 I mean, there's some level of strategic thinking, like trying to go beyond, trying to go beyond the zeitgeist, right? And figure out like what weird s\*\*\* may happen three or four years from now that no one's foreseeing, right? So, that, that sort of strategy thinking, I think also requires AGI. But the thing is, if you look at, you know, deep human contact, radical progress in art or science or wide-ranging deep strategic thinking, okay, these things probably do require a human-level AGI. On the other hand, what percent of the economy consists of these, of these things? Like, I mean, the first category of fundamental human contact is probably more of the economy than the other things. But I'd say 20% to be generous from all those things put, put together, right? So, if you think about it that way, it's like, okay, 80% of the world economy should be automated very soon, if not for frictions and rolling out technologies.

01:03:54 Now, frictions are significant though. Like in US, every McDonald's has some human being pushing the hamburger button. I mean, in Asia, most of them are, you go in to push the hamburger button on a tablet. But in US, we don't like that. We want the human pushing the hamburger button for us. I don't know why. I mean, it's the economics of McDonald's franchises or something, right? So, I mean that's a, that, that example illustrates

the frictions in rolling out relatively simple and mature technologies, right? Because for that, for that, you don't need any fancy machines. You don't need to automate flipping the burger or, or even sweeping the floor, right? It's really just pushing on the screen instead of waiting in line longer to have another human push on the screen. And pretty much everyone can navigate a menu to push the hamburger button, right?

01:04:47 I mean, so I mean that illustrates, even though 80% or so in my guess could be automated now, it's gonna be a while to all get actually automated. But I mean, there's, it's still gonna happen fast in many industries, right? And this is gonna have massively disruptive impact on the world economy. And I think in the developed world, this will inevitably lead towards some form of universal basic income. Because there's sort of nothing else to do. We're not gonna let half the US population like go homeless on the streets and rummage in the garbage. And in the developing world, it's just gonna be a lot more complicated because in say, sub-Saharan Africa, if most of the middle-class jobs go away, that just means more people go back to subsistence farming, which means they don't starve, but they can't pay their phone bill or get prescription medication and become very disgruntled at the world order that is, is pushing them backward while the rest, while the first world is like standing at home, playing video games and being served by robots.

01:06:05 And in the mid-level developing world, like say a Brazil or something where I, where I was, I was born. I'm a dual citizen in US and Brazil. So, I have some attachment there. I mean, it's too advanced for everyone to go back to subsistence farming. It's not rich enough to do universal basic income. A messy mix of things will happen. And you'll see a dynamic whereby the superpowers, you know, offer various unpleasant bargains to developing countries to help them with basic income. Right? I mean, so I think,

so the unfortunate ethical trade-off I see here is, the best way to avoid a bunch of horrible mess as the economy gets automated is to have rapid development of benevolent democratically governed AGI. On the other hand, the best way to ensure that the AGI we develop is benevolent is not to rush it and take time to be sure we've, we've got the motivational system and the ethical framework, right?

01:07:11 So, I don't, there's a trade-off here, which is very bad and I don't, I I'm, I'm actually an optimist about the end game, but, it seems like, there's complicated tradeoffs in, in the, in the next years as all this unfolds. And I mean, if you look at how the global economy works, like we can't, the WTO can't even agree on like IP for basic physical objects being manufactured. We can't f\*\*\*ing convince the superpower from not invading it's neighbors and senselessly blowing people up or convince the US not to fund like third world dictators slaughtering their population for, that matter, right? So, I mean, we're, we're very bad at regulating quite simple things. We can't, I mean, the music industry can't figure out how to pay musicians. Spotify just eats all, eats all the money, right?

01:08:10 So, very basic things we haven't managed to sort out, it's hard to see how we sort out all these more complex factors that are gonna happen when AGI is unfolding. I think the more we can make open-source, the more we can make democratically governed, the more we can make decentralized, at least you're giving interesting tools there that the world can use in ways that we can't currently prefigure in, in detail to, to cope with all these, all these things. But there's potential for a lot of mess, even if once you get to benevolent, decentralized like democratic AGI, it is, it is a rosy future.

Jon Krohn: 01:08:57 Yeah. I am with you on all of your points and you know, we've had systems that humans have developed but

quickly become out of our control for centuries or millennia, you know, market systems that, that are predominantly-

- Ben Goertzel: 01:09:13 Yeah, they become out of our control, yet they're within our control too, right? I mean, it's humans doing all this stuff, right? It's just their mind programs that are living in all of our brains and training us in a certain way. And that, I mean, with AGI also, like if you have a decentralized AGI running on millions of machines and hundreds of crypto mining farms in every country, I mean, until the thing has become superhuman, humans could just all pull the plug, right? The thing is that, the thing is that our collective thought systems get, get colonized by collective like mind viruses that, that direct us in the, in, in the, in a certain way. So, I mean, yeah. Yeah, we've lost control of corporations, but yet it's all people in the corporations making these decisions, right?
- Jon Krohn: 01:10:04 Right. Yeah, exactly. Yeah. So, these complex systems yeah, they're out of the control of any particular individuals. I guess they're kind of decentralized in that way, but yeah, markets, governments, armies-
- Ben Goertzel: 01:10:15 [crosstalk 01:10:16] like Cisco, which I've worked with off and on, feels like a decentralized autonomous organization. They've got hundreds of CTOs and no one person could list every product they make, right? It just, it's a great company. It accumulates money, it ingests other companies and assimilates their products and it's like a vast amoeba in the tech ecosystem, right? So, there, there's a lot of these sort of decentralized, autonomous organizations that don't rely on any one person and they just keep growing of their, of their own accord. I mean, Linux you would say is like that, which is right to my mind, a source for, for good predominantly, right? So, but, yeah. This, this is what happens. And as we go toward AGI, I mean, the question is, does it advance in a

way like the internet and Linux have, because these are the two big examples in my mind of sort of open, decentralized technology / human ecosystems that to my mind, by and large, have been a force for good and have evolved in a roughly democratic and decentralized way that with a lot of complex mess behind them, right?

01:11:36 Of course you can't say they're, they're all good. Some bad guys can use embedded Linux, you know, and in the OS for a bomb they used to blow up innocent, innocent people. I mean, al-Qaeda use internet to message back and forth. But on the whole, I feel the Linux and internet have been sources for good and they've been rolled out in a democratic decentralized way. So, I would like AGI to be more like the internet, then like the mobile ecosystem, and more like Linux, then like, you know, centralized operating system stacks and that, that doesn't seem entirely, entirely fanciful, although it's not how the AI ecosystem is evolving totally at this moment. But then you have, like, we all saw that there is No Moat paper leaked from Google, where someone in Google's AI development team was like, "hold on." The problem isn't us catching up with OpenAI. Sure we can catch up with OpenAI because we invented transformers in the first place. The problem is open-source will outpace all of us, and how can we stop that? We have no conceivable strategy for that, right? And I think, I think that was right and I'm very happy about that since I'm, I'm in the, in the open ecosystem instead of in the big company.

Jon Krohn: 01:13:01 Yeah, I think that's right as well. And actually two episodes ago we had a machine learning engineer, Lewis Tunstall from Hugging Face and Hugging Face is really big on this. They're also trying to build-

Ben Goertzel: 01:13:13 Hugging Face is amazing. Yeah, I love Hugging Face and the, they're the ones who made MusicGen, the music

model that I mentioned earlier. So, so easily, easily available also.

- Jon Krohn: 01:13:26 Yeah, they're really good at that, really quickly rushing to get any available new progress that they can open-sourced. And yeah, really driving things [crosstalk 01:13:35]
- Ben Goertzel: 01:13:35 Oh, it was great. I mean, I mean I had, I had to redeploy that myself to use it the way I wanted, but I mean, they had it there, you could play with it. I mean that's, that's, that's exactly the sort of thing we need, right?
- Jon Krohn: 01:13:47 Yeah, exactly. So, if I can squeeze in time for one last question, you have posited that a superhuman mind. So, if we realize AGI or, and then quickly on the heels of AGI, ASI that superhuman mind might perceive itself as a collection of patterns and subsystems rather than as a singular entity. And then it also seems like the way that you've been describing in this interview, as well as in other talks that you've given to the past, papers that you've written, that you and I, who, who primarily perceive ourselves because we have passports and driver's licenses today that say our name on them, and you know that the face in those passport photos looks roughly, roughly the same, even over a 10 year span. So, we kind of have this sense of, of an individual. It's even literally there in the name. I am an individual, you're an individual, we can't be divided, but in fact, we are just made up of biological material. Some of it comes and goes like the Ship of Theseus. And we also, like you mentioned, you know, the mind viruses, like we do have some of our own thoughts that happen to somehow materialize, but most of what we think and most of what we do is influenced by experiences that we've had and other thoughts that have infected our brains. And so, yeah, so the idea of me Jon Krohn, or you Ben Goertzel, as being this indivisible individual is nonsense.



- Ben Goertzel: 01:15:28 Yeah, I think, I think, I mean there's many possible kinds of intelligences, right? And there what kind of mind a certain AI system becomes has to do with what the AI system is, is doing, what experiences has. Now, our experience predominantly is controlling this, this body in, in the world. And so we naturally attach our experience to that. Then when you, when you talk to people who have meditated intensively for long periods of time, you find they often come to conceptualize themselves differently and they sort of, they don't center their experience around this sort of psychosocial self anymore. They experience, their experience is like there's this set of clusters of behavior patterns and then some of these behavior patterns have some, you know, models that they concoct for some temporary purpose associated with them. And that's, that's a different way of organizing experience than the way that most people have in their ordinary state of consciousness, right?
- 01:16:49 Now, if you're, if you're an AI system, you may land in that sort of state of consciousness right from the get-go. Because even if you're controlling a body like a Sophia robot or an avatar in SophiaVERSE, I mean, the same AI system can control a lot of the at once, or, I mean, you can save knowledge from one body, reload it in another body. And it can also do stuff like read the web or prove a billion math theorems that are unrelated to being stuck in a body, but they're just crunching data on some machine. So, it would seem likely that the most natural way for an AI system, if it's based, you know, in Azure compute cloud or on SingularityNET decentralized compute cloud, right? The most natural way for that AI to experience itself would be more like the, more like enlightened advanced humans.
- 01:17:45 More like just, okay, there's a set of clusters of behavior patterns that have different sorts of models and capabilities associated with. I mean, for humans to come



to look at it that way, we have to let go of a lot of ego and social conditioning. But for, for an AI look at it that way might be very natural just because the AI is not attached to a given body, like in the first place. And this is actually one of the reasons I'm optimistic on AI ethics. I mean, I think most people actually have a good sense of what is the ethical thing to do by standard human ethical judgment. And this is why ChatGPT is so good at answering ethics questions like "Given situation, what's the right thing to do?" It's very good at figuring that out.

01:18:35 And I think almost all humans are also, the problem with ethics isn't that humans don't know what our common sense says is ethical. The problem is that we would often rather do what's good for ourselves or, or our tribe instead of what we think is, is good for the whole. And I mean, me too in various points in my everyday life, I'm, I'm not a perfect hu human either. Now an AI system can have the same sense of what is ethical that we do. And you can see ChatGPT already embodies that sense in its own way, even though it's not a moral agent. It has a good ability to answer ethics questions, right? Then, but the AGI doesn't have to have a selfish or a tribal interest unless we program it and, and condition to it could come, it could come right out of the box with sort of general common sense of ethics as it as its main driver. We just have to configure it that way, right? I mean we're, we're defining what is the top level goal of the, of the AGI system. We evolved to fight to survive. And we have to work to overcome that.

01:19:43 Like, right, right now, if I'm in a really good mood, I take everything blissfully and calmly. If I'm in a bad mood and someone points out something I did that was wrong or stupid, like there's some anger that rises up inside me, I'm like, "f\*\*\* you. Why are you saying that?" I mean, I just tamp that down and let that like flow pass before it takes of me because I'm 56 years old, I'm not five years

old, right? But, but we all, we all have that, right? I mean, we evolve with that. You don't have to program that into the AI, right? We, we should not do so.

01:20:20 Now we could, if you want to build military robots and if you're building like a corporate sales AI, you may build into it that his motivation is "haha I figured out how to extract all this guy's money by selling him crap he doesn't need." I mean, you could build AI with that motivation, but we don't have to, we could build an AGI with a motivational system to like do, do what the average person would think is the most beneficial, loving, and compassionate thing in this situation. And if we program the AI to have that as its goal, it will do it and it will probably then evolve a quite, sort of enlightened, unattached inner life. Which, which is almost natural if you're a mind that's not by design attached to any one body infrastructure or hunk of matter, right? So, this gets into why I'm optimistic about the potential for beneficial AGI and why decentralized is important. Because centralization of control tends to bring with it some narrow motivational system separate from, from the ethics of what's best for everyone.

Jon Krohn: 01:21:37 Yeah, fantastic answer Ben. And I'm aligned with you on all of that. And it's nice to have a guest on the show where we can get into these kinds of topics. You know, I frequently have really brilliant researchers on the show and I ask them big questions like, you know, where's all this going? And, you know, what would you like to see happen in our lifetimes? Or, you know, what, what kind of world do you want your children or grandchildren to be living in? And it's pretty rare, it's surprisingly rare that people are willing that their minds are willing to unfold beyond six months, 12 months and try to figure out how things might look. So, this has been a fabulous conversation for me, no doubt for a lot of our audience as

well. Benn, before I let you drop off, I'm sure you'd have an amazing book recommendation for us.

- Ben Goertzel: 01:22:32 Let me think. I will I think I'll recommend a book from the late 1960s actually, which I read when I was a little kid, which is called The Prometheus Project by a Princeton physicist named Gerald Feinberg. I read this in 75 or so. He wrote it in 68. And what it said is, when the next few decades we're gonna get AI smarter than people, the ability to prolong human life indefinitely, and nanotech, the ability to manipulate matter at will, and our choice will be whether to deploy this technology for rampant useless consumerism or for expansion of human consciousness. And he proposed that the UN should put that to a vote of all global citizens as to whether to take these advancing technologies and direct them toward consumerism or toward consciousness expansion. And that, it's funny now to look back at how this was, was conceived like back in the, in the late 1960s because it's, it's sort of all there but expressed in an, in an archaic language.
- Jon Krohn: 01:23:47 Very cool. Love that recommendation. And obviously, anyone can learn a ton from you as they did in today's episode. After this episode, how can they follow you?
- Ben Goertzel: 01:23:58 Yeah, great question. So, SingularityNET.io has links to the various different media outlets associated with SingularityNET of YouTube and Twitter and blog and so forth. For me personally, if you can look at my website goertzel.org, G O E R T Z E L .org or follow me on the Twitter or YouTube where I'm just Ben Goertzel and lots of stuff coming out all the time.
- Jon Krohn: 01:24:30 Nice. All right Ben, thank you so much for taking the time with us today and yeah, really mind-expanding episode. Thank you so much.



- Ben Goertzel: 01:24:38 Yeah, yeah, thanks a lot. Great, great conversation.
- Jon Krohn: 01:24:46 Whoa, so much to reflect on from today's episode. In it Ben filled us in on how benevolent decentralized AGI would not have a single owner and be for the benefit of sentient beings. Talked about how neurosymbolic systems, genetic algorithms, and knowledge graphs could be combined with deep learning and potentially realize AGI on a startling three to seven-year time span. He talked about how virtual or perhaps even physical interactions between Sophia Humanoid robots could also potentially give rise to AGI-level intelligence through unstructured exploration. He talked about how the capabilities of AI today are not far from being able to replace, in his view, four out of five paid jobs that humans do. He talked about how if we realize AGI, the artificial super intelligence, and therefore singularity that it could immediately give rise to, will dramatically transform society, including by perhaps giving way to an interconnected hivemind and things like femto-scale modelers that could then print any desired object. And he talked about how like a master meditator, an AGI or ASI may be aligned with the best morals of humans and unattached to outcomes.
- 01:25:52 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 Dr. Goertzel's social media profiles, as well as my own social media profiles at [superdatascience.com/697](http://superdatascience.com/697). That's [superdatascience.com/697](http://superdatascience.com/697). And if you enjoyed this episode, nothing's more valuable to me than if you take a few seconds to rate the show on your favorite podcasting app or give the video a thumbs up on the SuperDataScience YouTube channel. And of course, if you have friends or colleagues that would love the show, let them know.

Show Notes: <http://www.superdatascience.com/697>



- 01:26:24 All right, thanks to my colleagues at Nebula for supporting me while I create content like this SuperDataScience episode for you. And thanks of course to Ivana, Mario, Natalie, Serg, Sylvia, Zara, and Kirill on the SuperDataScience team for producing another mind-bending episode for us today. For enabling that super team to create this free podcast for you, we're deeply grateful to our sponsors. Please consider supporting the show by checking out our sponsors' links, which you can find in the show notes.
- 01:26:48 And finally, thanks of course to you for listening. I'm so grateful to have you tuning in and I hope I can continue to make episodes you love for years and years to come. Well, until next time, my friend, keep on rocking out there and I'm looking forward to enjoying another round of the SuperDataScience podcast with you very soon.