**Anna:** 0:00

So customer comes to us, they've got a spec, and again this is across all different industries, so it's as varied, as I want my corn to fertilize itself and I'm interested in engineering a microbe that will live inside the corn and fix nitrogen out of the air and so I can reduce my nitrogen addition by 50% while not harming plant yield. That could be the spec, or the spec could be against something as specific, as I said, catalyzing a particular reaction, or it could be. I want to discover a new RNA therapeutic.

**Craig:** 0:33

Hi, I'm Craig Smith and this is Eye on AI. In this episode, we explore the dynamic world of synthetic biology with Anna Marie Wagner, head of AI at Ginkgo Bioworks. Anna Marie shares how their platform leverages AI to push the boundaries of biotechnology, touching on projects like self-fertilizing corn and the quest for biodegradable materials. We delve into Ginkgo's collaboration with Google, the nuances of building foundation models for biological data and the broader implications for biosecurity and sustainability. The conversation provides a window into a future where biology and machine learning converge to innovate and heal. I hope you find the conversation as fascinating as I did. Introduce yourself, anna Marie. Tell us a little bit about who you are and where you came from, and what you do.

**Anna:** 1:37

Sure, thanks for having me on. So I have been at Ginkgo for about five years. I wear a number of different hats. I am our head of corporate development and so in that role I oversee M&A licensing, strategic partnerships, and that was the role I joined to lead. But as part of that, I negotiated our collaboration with Google, which I'm sure we'll talk about later, and started working very closely with all of our scientists across the business that are deploying AI across many different disciplines and got, candidly, just really excited about everything that we were doing in AI and the potential for AI in our field, and so took over a new role about five months ago or so as our head of AI, overseeing our AI strategy and commercialization efforts, so really working with our customers.

**Craig:** 2:25

Yeah, and so tell us about Ginkgo Bioworks and what you guys are focused on. I mean, there's so much happening in biotech and AI.

**Anna:** 2:39

Yeah, absolutely yeah. So it might be helpful if I give you just a little bit of a history, not just of the company but of the backgrounds of our founders, because it does help elucidate some of the philosophy by which we've approached the business. So Ginkgo itself was started about 15 years ago, five co-founders, all of whom are still with the business now, 15 years later, but they've actually been working with each other for over 20 years. So they all met at MIT. Four of them were doing PhD programs in the newly created sort of bioengineering field or kind of synthetic biology field, and the fifth was a professor. His name is Tom Knight. He had been at MIT for 48 years, originally in the electrical engineering department. So he built one of the early semiconductors, worked on LISP, built the first real-time debugger. Eventually, as semiconductors came onto the scene, he taught the semiconductor design class at MIT and it was around the early 90s. He got interested in biology for a couple of reasons. One is as a substrate for nanoscale precision, which has a semiconductor designer, was something he thought a lot about as he thought about the future of Moore's Law. But also because biology is a programmable physical discipline and so it runs on code ACTG, not zeros and ones. But we could read that code with DNA sequencing, we can write that code with DNA synthesis and he's like we should be able to program this stuff. And so he's approaching the field. Well, what would it mean to program the physical world? That's pretty neat. And so, to his great credit, he's in his mid-40s. He starts taking undergrad biology classes, he builds a wet lab in the computer science department which, as you can imagine, the rest of the computer science department was really thrilled about and he started assembling all of these engineers from different disciplines mechanical engineering, chemical engineering, computer science, electrical engineering, obviously to start thinking about what would it mean to make biology an engineering discipline rather than just this artistic field of discovering nature? So a very different way of thinking about the field. And there were a few observations relatively early on. One was something we needed to build first of all was the layers of abstraction. So to be a programmer of biology, the first thing you were told to do is pick up a pipette. It was sort of like being a programmer of computers in the 60s when you program a computer, you needed to understand the electronics, not just programming languages to exist, and so we are still living in that era of programming biology where to be a programmer you had to understand the fundamental mechanics of the machine. And so he's like, ok, how do we build those layers of abstraction? That was kind of one key question. Another key observation was because this is a physical discipline, physical disciplines tend to benefit from scale economics. And so how do you have scale economics in biology? You're going to invest in the sorts of things that drove scale economics and things like industrial manufacturing, right? So automation, miniaturization, things like that, and that led to business model decisions. Ultimately, 15 years later, when Ginkgo got started around, being a platform that was going to serve many different areas of biological design rather than focusing on a single product, which wouldn't allow us to make the kinds of investments at scale that we knew were going to be needed in order to actually drop down the cost of doing this engineering work sufficiently to make it scalable. And then the third piece and final piece that he really noticed or question he asked at the time was well, where are the code libraries? I'm going to programs in biological code when is the library? Like we have this stuff in computer science, where is it in biology? It didn't exist. All of the code that had been created was happening in silos, in small biotech companies that were making their individual drug, and they're religiously protective of it, because the only thing that creates value in biotech is IP, so everyone is protecting their IP, and so there was a ton of recreating the wheel that was happening in the biotech field, and so that again led Ginkgo to make a very different kind of business model decision, which is that when we create IP, we want that IP to be shared broadly. We know that if we engineer a protein that's useful for one domain, it's very likely useful in many, many other domains, and we've seen that time and time again in the history of Ginkgo. But that was quite counter-cultural in the field, and so I just want to give that history, because it helps explain what Ginkgo is today, which is a horizontal platform with a very large 300,000 square feet or so of infrastructure, largely automated labs, that is acting as a platform to enable R&D to happen more efficiently, faster, ideally higher probability of success for customers across many, many different industries. And then the business model alongside, that is, we get paid sort of like AWS service fees for doing the work on the platform and then more like an Apple App Store or something we're getting paid on success. So if the customer's program actually works and they sell product, we get a royalty on that.

**Craig:** 8:05

And what kinds of products? Because? Is this drug discovery, Is this engine gene editing of plants? Is it? Yeah, sort of give me a scope of what you guys covered.

**Anna:** 8:21

Yes, so, yes, all of the above. So, sort of counterintuitively, we got our start outside of pharma, even though it's obvious to everybody that pharma is the biggest market for biotech. Had we walked into Pfizer 15 years ago and we were like five people and some robots, they would have laughed us out of the room. They were way better at doing biotech R&D than we were 15 years ago, right, and so we started in areas like flavors and fragrances clearly a biological product, right, these are specialty compounds that are extracted from flowers and fruits and things like that. But the companies that were making these things had no real biotech expertise. They had supply chain problems, they had cost of goods problems and they didn't really know how to solve those. And so it was a nice interface between our capability set and our partner's capability set, and so that was where we had our first product. So, companies like Robertay and Nijivadan very large European flavors and fragrances companies where we got our start, and then, sort of slowly, we marched up the credibility curve where we then started working in more and more sophisticated end markets because we had the data to show. All right, we successfully delivered on products and this air programs in this area. Let's go tackle the next area. So specialty chemicals, food, agriculture, ultimately pharma. Today our business is split about a third, a third a third between biopharma, so customers like Pfizer, merck, novinortis, biogen all customers of ours across many different modalities in biopharma and everything from discovery to manufacturing. And then in the chemicals industry, we work with Sumitomo, we work with Solveys, some of the big chemicals companies. And then in agriculture we work with Bayer, syngenta and Corteva, three of the biggest ag companies in the world, both on things like traits but also on things like microbes that live in the soil and help fertilize your plants for you in the soil. So really interesting variety of work that is running on a common platform, which, again, is kind of unheard of in biotech land.

**Craig:** 10:30

Yeah, that's fascinating. And just to clarify my understanding because this is way out of my area of knowledge In pharma, there's a lot of talk right now about small molecules as opposed to biologics. Are you guys focused on biologics or does small molecules? Is that part of the mix?

**Anna:** 10:58

We do everything. So, yeah, we have small molecule programs and we have complex. So everything from small molecules to large molecules to complex therapies like cell and gene therapies, we have programs in On the small molecule side. Typically, where we have a particular advantage is when you're looking at what are called natural products. So these are small molecules that have otherwise been created in nature. So, for example, a lot of our antibiotics come from microbes that live in the soil because all these little bacteria are fighting each other, and so if you wanna fight that bacteria, turns out some other bacteria is probably already successfully fought it away. And so we found a lot of that chemistry by looking at biology. And what's interesting is that quite often these sort of magical molecules are quite hard to synthesize chemically. They're very, they're quite complex chemistries, and so even though they're small molecules, making them with traditional chemical synthesis pathways is hard, and so you also need biology to produce them. So that's one version and then more on the manufacturing side. Similarly so, for example, merck is one of our customers in this space. They're producing small molecules. They already sell these drugs. Right now. They might have some really expensive or messy or inefficient chemical synthesis stuff, and so one of the project areas we work on is something called bio-catalysis, which is can you do a chemical synthesis step using a protein, using an enzyme? And so we would discover, an, engineer, an enzyme to catalyze that chemical reaction more efficiently than the current chemical synthesis pathway that they're using. So those are a couple examples of small molecule products, but the industry writ large is quite focused on biologics and cell and gene therapy as well.

**Craig:** 12:48

Yeah, where does AI come in? To the process, to your workflow?

**Anna:** 12:54

So again I'll give you a little bit of history. Ginkgo, historically, I'd say for the first 14 years of our 15 year history. Ginkgo is sort of accused of being this brute force experimentalist, like we're just gonna solve all the biological problems by throwing scale at it. And remember, biology is this like artistic discipline, so really good biologist sort of looks down their nose at this brute force scale that we were applying. And, to be honest, the way that Ginkgo originally thought about data and remember that those code libraries was around, the successful experiments, like we're gonna reuse debugged code that we know works, we're gonna reuse those across applications. But, as it turns out, the real magic of operating at the scale that we have operated at and that we've built is that we are not just studying the success stories, we're also collecting a whole lot of data on the 99% of experiments that did not generate success. And, as you well know, when you're thinking about training a model, you wanna have, yes, examples of what good looks like, but you also wanna have a hell of a lot of examples about what bad looks like, and you can learn just as much from that, if not much more, just because there's definitely a lot more failures than successes in biology. And so now, as we think about what is our role in AI as applied to biology, the rest of the industry really hasn't been designed or set up in a way to collect that kind of data real labeled data on how biology functions at scale Like that is a pretty unique infrastructure that we've built. So, to give you a couple of tactical examples, our protein engineering team big users of AI, so we've deployed several AI models inside Ginkgo and sort of homegrown, and we will use anything that's out there. We're sort of ambivalent as to what the architecture is, because the thing that we really bring to bear is data. We have a much, much larger genome collection than is publicly available. Our proprietary metagenomic collection is about 10 times the size of the public databases. And then we have all this labeled data on how those proteins actually function in the lab against a battery of tests. And so now when we're designing a protein for a customer, we can deploy all that data into these models and fine tune them to yield better results for the customers. So protein engineering is a big area for us. Dna design is another big area for us. So, for example, trying to improve the expression of a particular gene sequence by engineering promoters and other kind of non-coding regions of DNA. Those are kind of easy use cases for us today, but the reality is, anywhere we're generating data, that data is useful for training increasingly complex models over time, and so you can certainly imagine this getting to a point where we're creating models about how an entire cell functions. Right now we're starting with the building blocks proteins, dna, rna then you get into things like systems, biology and pathways and things like that. Then you get into a broader cellular function. Then you could imagine trying to predict how entire ecosystems behave together. There's an immense amount of complexity that is beyond the scope of what anyone can really do right now. But even the advances we're seeing just on those building blocks are quite remarkable.

**Craig:** 16:20

And so the AI that you're working with is primarily a search function. You're looking for attributes in a database of molecules, or are you using it to generate new molecules that have certain properties?

**Anna:** 16:46

Yeah, so it'll be both. So again, let's take one of those enzyme engineering projects that would typically start with an ML guided search campaign of the metagenomic landscape that we have to identify interesting starting points, and then we would use those starting points, generate data about them and then, in a kind of a closed loop system, generate new designs that we think are interesting. And what's important about the way that we operate again, because we have the labs and we can actually generate this data at scale in a pretty quick way is, rather than on the first round, saying, all right, we got one shot, let's make the best thing we can make. We can really design the experiments to maximize learning, and typically we're optimizing across many different parameters at the same time. So we will generatively design new proteins, which are going to be based on a scaffold that we've found in our databases from an existing piece of biology. And that's useful. Because you don't want to. There's a reason all that biology evolved and usually it evolved because it functions well, it grows well, it's soluble. There are going to be a bunch of features of the biology that's useful and you want to preserve. So we'll start with that ML guided search. We'll then generate a library that maximizes our ability to learn across the many different parameters we care about. So it might be activity, it might be selectivity, so fewer off target interactions, and then we would design subsequent libraries to maximize across those dimensions in parallel. But those are also using generative models to create new protein designs.

**Craig:** 18:26

In that case, yeah, do you know the company in Silico, which is a drug discovery company? They're using AI to generate or discover new molecules small molecules but they also have an automated wet lab in which they can synthesize these molecules, or at least samples, and then they send it out to a larger lab for production. Is that similar to what you guys are doing, other than the fact that they're focused on developing drugs and you guys are really a platform for anybody to use?

**Anna:** 19:11

So I think there are a lot of, there's a lot of really great research that's happening across that intersection of AI and biology. I think what's interesting about Ginkgo is that, again, we're quite ambivalent as to what architectures are going to end up working best, and we tend to be real beneficiaries of that kind of innovation that's happening elsewhere in the industry and, interestingly, I honestly see more opportunities to partner than to compete, and so we can in many ways operate in a couple different venues with some of these AI focused companies. One is to help their platform reach more customers. We just integrate the model. Today, our protein engineering team uses seven or eight different models in all of our campaigns, so we'll integrate different models into our work and we'll lean on the ones that then start yielding the best results, and we're obviously building our own internal models, but there's benefit to having that diversity. And then the second is, as you've seen, in the large language model space with human language specifically, there are companies that have specialized in building architectures and companies that have specialized in building data, and so take, for example, openai building GPT-4, but then ScaleAI, which created a lot of training data and allowed them to do reinforcement learning, so that you got chat GPT out of it. One way to think about Ginkgo is that we're like a ScaleAI for the broader kind of biotech ecosystem as it relates to AI. We're creating that data that's going to make all of these models better, and so our goal is to get a customer result. We don't have our own pipeline or anything else, and so we're very happy to use the best of what other folks are developing, and where folks have labs, they tend to be quite narrowly focused on the questions that they're interested in for their own model or their own pipeline, and so there do tend to be still really great collaboration areas. To round out the type of data that you would want, or, to the point you made earlier, once you start thinking about scaling this up, making it more industrial, wanting to really manufacture this stuff, how do you do that?

**Craig:** 21:21

And can you walk me through an interesting, because I'm sure you have many case studies at hand an interesting case study of sort of what happens with the client when they come to you guys, what's the process and then the workflow and the work product is yeah.

**Anna:** 21:50

Yeah, it is quite interesting because we're often compared to a contract research organization, a CRO, and in many ways we are like a CRO. But traditionally the way a customer would interact with a CRO is they would say I need this study done. Can you please do this study for me? It's an outsourced service, but the customer knows the work that they want to get done. When customers come to Ginkgo, certainly we're providing them a lot of services, but they're not dictating. All right, I want you to run this assay on your math specs and I want you to run whatever this fermentation process in your labs, etc. They're coming to us with a problem. I need to replace this chemical step with an enzyme and it needs to cost less than this. It needs to catalyze the sufficiency. It needs to have this level of specificity. Like they're giving us specs of a product, they're not giving us a research plan. We develop the research plan and it quite often looks very different from what the customer either might have tried internally or what they would be able to do anywhere else. Because of the scale of the platform we've built, we're able to try a lot more breadth and quite often what we see is you're trading a local optima for a global optima right. When you can only afford to search a relatively narrow space, you have to make pretty safe bets. You can't afford to try off-the-wall things. You can increase the scale you're operating at by 3, 4, 5, 6 times. Then you can suddenly start to ask much more interesting questions. We very often see that the winning design for a particular program might look absolutely nothing like their best design coming into the program. Customer comes to us, they've got a spec. And again this is across all different industries, so it's as varied. As I want my corn to fertilize itself, I'm interested in engineering a microbe that will live inside the corn and fix nitrogen out of the air. I can reduce my nitrogen addition by 50% while not harming plant yield. That could be the spec. Or the spec could be, again, something as specific as I said, catalyzing a particular reaction, or it could be. I want to design. I want to discover a new RNA therapeutic. I don't even know what the right design should be. I don't know if it should be circular, I don't know if it should be linear. I don't know how to deliver it. You've got to work on a whole package. For some of these projects the customer has got a product idea in mind, and then it's our scientist's job to bring together the pieces of technology that are necessary to deliver on that.

**Craig:** 24:33

Yeah, actually, of those examples, the corn is the easiest one for a layperson to understand. Was that a real example, that's?

**Anna:** 24:43

a real program we're working on with Bayer?

**Craig:** 24:45

Yeah, so how do you start and what's the process that you go through?

**Anna:** 24:55

Yeah, so that is it. That is a real moonshot problem. It's like a $70 billion a year industry in nitrogen fertilizer, so it's a really cool problem to work on. It's also, by the way, fun fact. Nitrogen fertilizer alone is something like 5% each of global energy consumption and global greenhouse gas production. So I mean, this is like if you can reduce the amount of nitrogen fertilizer that is needed to feed our planet, like this is hugely, hugely, hugely valuable. Okay, so where would you start? So that the there are sort of two components to that. This is a drastic oversimplification, but simplistically, there are two components to this challenge. One component is how do I get the bug to live happily inside corn or wheat or rice, or you know, you name it your cereal crop of choice. And so there, your engineer, you're trying to find bugs that live, and there already are, by the way, lots of little microbes that live inside plants very happily. And then the second key question is how do you get that bug to fix nitrogen? Well, it turns out there are bugs in the world that do that naturally, like you don't need to fertilize soy. Like crop rotation you might remember from like middle school we used to rotate crops and the reason you rotate crops is some of those crops, the legumes. They fix nitrogen, they replenish the soil with all of those nutrients because they've got these little microbes that live in their root structures that fix nitrogen. They do Haberbosch, basically. And so now the question is can I figure out the engineering circuitry of the bugs that do Haberbosch and can I put that circuitry inside the types of bugs that are very happy living inside corn or inside rice? And so you're and I suppose you could do it either way, but like you could try to make those bugs be happier with corn. You know, there are obviously many different ways to solve the type of problem. But you're figuring out what is the DNA that codes for Haber-Mosh, basically, and then you're trying to figure out what is the DNA that codes for corn lovingness, and you're trying to get both of those things to exist in the same bug. And then you've got all sorts of things like how do you deliver it? Is it a seed treatment? Is it a soil additive? You know all sorts of other things, but that would be the basics of that type of a program.

**Craig:** 27:20

Right, so let's take the, which sounds simpler, or maybe not, but figuring out the genetic mechanism for fixing nitrogen in the soil. Yeah, it walked me through how theoretically, how you would do that.

**Anna:** 27:45

yeah, yeah, so it's a metabolic pathway, right? So the microbe is eating something in the air, right, and it's like what's in the air? There's some nitrogen floating around, there's oxygen floating around, there's some carbon floating around, right, and these bugs are ingesting those molecules. And then there are a series of proteins that are coded in the DNA, and those proteins catalyze certain chemical reactions, and so many of those chemical reactions we sort of understand, like we understand generally. Well, first you know, all right, the plant eats the carbon dioxide and then the carbon sorry, I'm not a biologist anymore, I don't even remember these numbers the carbon dioxide is broken down into something else which all you know. Like we do understand these metabolic pathways in general, and so then we can start identifying the proteins that are responsible for catalyzing those reactions, and then it's a question of so you generally have a decent starting point as to what the set of chemical reactions needs to be. Sometimes, by the way, you can find more efficient pathways. So, like, hey, what if I took out these two steps and instead just did that one, that kind of circumnavigated the chemical? This is chemical steps, but typically you'll find, okay, there, I don't know, let's say, four steps involved. I'm making this up four steps involved in converting, you know, carbon dioxide plus you know whatever else into ammonia effectively. Here are what those steps are. Here are the proteins that code for those in the microbes that live near soy. Then there's a question of, like you know, load balancing, basically like, how much of these things do I need to get expressed? And so that's where understanding things like promoters, which are like non-coding regions of DNA, matters, because you need to make sure the right amounts of these proteins are getting expressed in your new bug. You need to make sure your new bug isn't getting killed because of these new pieces of DNA that you're adding or any of the intermediate molecules that are floating around in the cell now that are being catalyzed. So the hard part is not typically let me figure out the proteins that are responsible for this, it's let me figure out how to engineer those proteins, that those pathways that make those proteins and catalyze those reactions, into an organism that has never had to do that kind of work before. Because you might then find, oh well, it turns out that this intermediate compound is building up in the bug and the bug is dying, and so you're not ever making the ammonia, and so those are the so then you might have to engineer a bug that is more resistant to having that compound built up in it, for example. We have this all the time, actually, where, like in the industrial chemical space, many of the products we need to make are acidic, and a lot of bugs aren't very happy living in very acidic conditions, and so then we need to actually engineer the bugs, not just to make more of the acidic thing, but to be able to survive in this broth of acid that they're now living in when they're successfully making that. And so you end up again with these really complicated multi-parameter optimizations, and many times going in, you don't even know how many parameters you're going to ultimately need to optimize over. It's why they're hard problems and it's why, like AI is so important. It is like the human mind sometimes just can't even comprehend all this kind of the amount of data that we are generating, and to be able to use some of these tools to find bits of signal and all that noise is really quite critical.

**Craig:** 31:12

And then so you're using AI on this tremendous amount of data to find an optimal solution. Once you found the optimal solution, then what's the next step? Or is that, as are you done with your work? Then it depends a bit on the customer.

**Anna:** 31:34

So some of our customers have their own internal manufacturing and they're quite sophisticated as it comes to scale up, but it's not just about the amount of data that's being generated. They're sophisticated as it comes to scale up and further development, and so we would just send them a tech transfer package that would probably have the organism and some instructions on how to manufacture and things like that. We'd give it off to them. Some of our customers have no ability to manufacture biological products themselves, and so we would partner on their behalf with a contract manufacturer and do a similar thing Tech transfer the organism with the manufacturing conditions. That tends to be. Basically, the final package is an organism that produces whatever it is that the customer is interested in, along with, effectively, a recipe for how to manufacture it.

**Craig:** 32:26

And there's a lot of concern about genetic engineering because it's not clear what the implications might be Implications might be, once organisms are released into the wild and start interacting with other organisms. How do you? I mean, presumably safety is something that's part of your remit how do you? Is that again? Are you using AI to explore, to sort of look forward, to explore potential outcomes as an organism interacts with other elements in the environment? How do you deal with that?

**Anna:** 33:19

Definitely a major focus, certainly in the agricultural field. To that point, if you're releasing something into the wild, there is a lot of focus on can you control it? The good news, to set everyone at ease, is usually the problem is the opposite. Right, like your bug dies because you've just engineered it to be really metabolically inefficient. Right, it's busy making a bunch of nitrogen, not busy reproducing and surviving. So usually the issue is the inverse. But yes, we obviously spend a lot of time thinking about bio safety and, specifically, we've thought a lot about bio security even well before the pandemic, where suddenly everyone started caring about the risks of biology. Our view was that it's, whether it's nature made or man made, we are made of biology and so there's no question that, and we eat biology and most of our materials come from biology. There's no question that biology is impactful, but there's also no question that we are very vulnerable to biology. Our food system is vulnerable to biology, our bodies are vulnerable to biology, our environment is vulnerable to biology, and so it is somewhat preposterous, candidly, that we do not today have a good understanding of what biology is floating around us at any given moment. We have radar for the weather. We don't have radar for what pathogens are floating around my kids' school, and I'm sure there are a lot. By the way, right, had we had the equivalent of radar for biology, you know, four years ago, would COVID have spread as far as it did? Maybe not, like, maybe you could have responded before people started dying in hospitals Like we. Really. That should have been a wake up call for us to start wondering about this. What's happening is at the intersection of AI and biology. There's a lot of fear, right, like. Is this intersection going to create a moment where suddenly it is easy to create bio weapons or create products that are so powerful that we can't control them, et cetera, et cetera, et cetera, and in general, again I would like, biology is still really hard. We can all relax a little bit, but I do think that it is not enough to rely on things like red teaming and building safer models and relying on good scientists to do good science, like we have to build an infrastructure to help us figure out when things have gone wrong Again, whether those things are manmade or nature created. Like there's a lot of biology out in the world that humans haven't touched most of it, in fact, that is still very, very, very dangerous to us, and so we really need to start building up an infrastructure to detect and respond to biological threats. And some of the work that we've done, like with AIARPA, which is basically the DARPA for the intelligence agency, and with the Defense Department, is around this type of thing. So like if you find a new sequence of DNA, can you tell, even if it's manmade or if it was naturally occurring? Like it actually matters to be able to answer that question right. Can you figure out what it does? Is it something you should be worried about or is it totally harmless? Like these are important questions that we need to be able to answer and again, today biology is kind of a big black box to everybody. This is cutting edge technology in very much an area where we've deployed AI and I can go.

**Craig:** 36:47

Are you working on new materials? Do any of your clients, or any of them, looking at new materials? I mean, you know, plastics is an issue that people are concerned about the proliferation of plastics in the environment, and I keep expecting these. I've talked to a lot of people in new materials labs who say that there are solutions being developed for biodegradable plastics or microbes that eat plastics and that sort of thing, but it doesn't seem to have arrived in commercially. So do you do any work on that?

**Anna:** 37:50

We do, yeah, and I've got a lot of thoughts on this topic, I think so maybe a philosophical lens or maybe just a framing lens on this, like I do like asking the question like what can't biology do? And it's a very small list. Like plastics are obviously something biology can do. They are made today with fossil fuels. Like fossils, those are biology. Right, it's dead biology Over billions of years. It turns into oil. We make petrochemicals, ie plastics, out of it, so it is very clearly a biological product. There are lots of companies today working on being able to replace traditional petrochemicals with biologically derived versions of those petrochemicals, and they tend to be biodegradable versions of those chemicals, of those chemistries. The issue has historically been cost. We are not pricing in the externalities of extracting oil into the price of oil, and so this has either a social solution or an economic solution, right? So? Or, as I should say, a technical solution. So an economic solution or a technical solution? The economic solution is the price. If the price of oil is higher, then a lot of technologies that are already on the market suddenly become economic, okay. The second solution is technical, which is like today, most of these fermentation processes that are creating biomaterials are relatively basic. You've got a big 50,000, 200,000 liter stainless steel tank. You got a bunch of bugs in there. You're feeding them sugar and out comes a petrochemical. Well, if sugar is more expensive than oil, you are never going to get a petrochemical, a fermented chemical, that is cheaper than a synthetically derived chemical. So there's where, like, a technical innovation can really solve it right. Well, what if we start feeding these bugs carbon dioxide instead of sugar? All you need to do is figure out how to get these things carbon, and you can make a lot of different products with it. It's just turns out that's not an efficient process. Today it's like it's a pretty hard technical challenge, but there are a lot of folks working on that one too, and so I suspect that both of these will advance kind of in parallel, where it's like we will start more and more and more realizing the need to have alternate solutions beyond current oil derived products, and the technology is advancing now in a clip where we will probably make increasingly make biologically derived commodities that are cheaper than their synthetic counterpart, and that will be a really interesting moment when that occurs.

**Craig:** 40:27

Yeah, and from your position in this world, you know, from the layman's point of view, things seem to be moving very fast, at least on the research end. I mean, I've been paying attention to machine learning for a while and I could see that, you know, from the advent of deep learning and then the transform algorithm that this was gonna change everything. And the people around me I would tell them this is, it's gonna be amazing, and everyone kind of in my family sort of humored me but weren't that interested. And now you know everybody is talking about AI. Do you feel that way in computational biology or synthetic biology or biotech or however you wanna refer to it, that there's a lot of stuff happening that hasn't quite hit the public sphere yet and the day will come soon when everyone is talking about it?

**Anna:** 42:01

I mean counterpoint. We all know what PCR is now. We all know what RNA is. You know we've all been gene edited for the most part at this point, right, because we've got the COVID vaccine, you know. So I do think that COVID created like silver lining of COVID, like it did create an understanding of both biological risk and the opportunities that come from cutting edge biological research, and so I do think it has changed. I think we're behind AI, obviously, and I would say candidly, like it is starting to change now at a fundamental level, not just a perception level. But for the past 40 years, biotech innovation has moved slower than I think any of us hoped it would have. You know, certainly has been impactful, but probably could have been a lot more impactful if the industry was organized a little bit differently, had slightly different incentives and had better tooling and stuff. And I think what's happening now is the intersection of a technological shift and a cultural shift in the industry, which I do think is going to allow innovation to progress quite a bit faster than it has over the past 40 years. Yeah, 40 or so years.

**Craig:** 43:18

Yeah, and for Ginkgo Labs. Or do you call yourself a lab, I guess, we call it the soon.

**Anna:** 43:28

We call our lab the foundry. We borrowed the language from the semiconductor industry. But yes, they're wet labs, very large wet labs.

**Craig:** 43:37

Is that kind of platform scalable to the point that there'll be? It'll accelerate progress? Or is it still such capital intensive and expertise intensive enterprise that there'll be these specialized companies like Ginkgo Bioworks? How will this scale?

**Anna:** 44:07

Yeah. So look, I think it's both. Yeah, I think we have certainly created a level of scalability that is unheard of historically, has been unheard of in the space. The challenge is twofold. One is decoupling kind of scale of research from humans, and so that's sort of the first step, and we've made a lot of progress there. But most labs scale with the number of PhDs they hire. That is not scalable. They're excellent PhDs but you're hiring people to do manual. You're hiring highly trained scientists to spend most of their time moving clear liquids around a lab. That is crazy, but that is what 95 plus percent of research looks like today. So Ginkgo has largely abstracted away that again, that physical process of programming, which now our robots do, from the scientific process of designing a program. The second step then is how do you scale the underlying infrastructure that's doing the programming? And so this maybe the analogy I would give is like when you go from, you know, vacuum I'm not an electrical engineer but like vacuum technologies to like microprocessors or something like, we are still using these relatively crude robotic devices to do that work. Now, that's way more scalable than human, but it's still like a physical thing that is reasonably, you know, clunky. Many of our most advanced processes. Right now we're using biology to solve the engineering process right. So, like biology, again, it operates at nano scale. It's really quite an efficient little machine, and so we can sometimes create a biological assay that abstracts away the need to do a lot of the physical experimentation. So, for example, pooling millions and millions of designs into a single well of a plate where every single design has a barcode written in DNA, and so then, instead of having to do a million different wells on plates, you have one well with a million designs, and then you can figure out which design was the one you liked by sequencing the barcode. You know. That's where you get like remarkable step changes in your throughput. Now, today there is still a trade off between the kind of quality or depth of the data that you're going to get out of a pooled process like that versus the arrayed format. That's more traditional, but those sorts of advances are the things that I think again get us the step changes in scale that are quite useful.

**Craig:** 46:46

Yeah, and what's the collaboration with Google?

**Anna:** 46:50

Yeah, so we have a nearly $300 million cloud collaboration with Google, so they are certainly our preferred cloud provider for building and training our AI models. But then what was interesting about collaborating with Google is that they really saw Ginko as an ecosystem enabler. So if they go make a great cloud deal with pick your favorite biopharma company, that doesn't help them go get a deal with some other big biopharma company other than maybe it's a logo on a page. That's a nice reference. You know that's about it. The way they looked at Ginko is, if Ginko is building foundation models for AI, for AI, foundation models for biopharma, then and if we're successful in that, then that helps bring the rest of the industry on to Google and into using more of these computational tools Because, candidly, it's just not that big a cloud market today and it should one day be the biggest cloud market, I would argue and so they're funding over about $50 million of basically R&D and model building at Ginko for us to build foundation models with, so build the team, do the training, et cetera, which was neat.

**Craig:** 48:06

So you are building foundation models on Google infrastructure and from scratch. Are you fine tuning models that Google provides, or open source models or something like that?

**Anna:** 48:25

Yeah. So, again, our goal is to deliver the best technology to our customer, so we are students of the architectures that are developed elsewhere in academia, that are released into the open source community, that other companies have, and even today we have brought many of those models into Ginko, and we have fine tuned those models using our data, whether that's labeled data in a specific domain where we want to answer a very specific question, or broader, unlabeled data sets that are minimally labeled, data sets that help understand a wider protein space. So I would say yes and so we are building foundation models for proteins, dna and RNA, but what we end up deploying to the customer is typically a fine tuned model that is relevant for their particular question or particular area of focus. That would be trained on the labeled data that we're generating in the labs. So both Our theory, though, is that those fine tuned obviously those fine tuned models are a lot better if they are trained on top of a stronger foundation model, in the same way that chat GBT got a lot better when GBT4 came out. You know, we think those foundations should evolve, and we want to be building our fine tuned applications on top of the best foundations possible and the theory, there being data is the missing link right now in building strong AI models, and that's what we have a lot of. So we'll be students of architectures that other folks are developing and practitioners when it comes to ingesting a whole heck of a lot more data into those models.

**Craig:** 50:02

Yeah, and maybe I'm I'm wrong, but it just it seems that a building, a foundation model is is an entirely different domain, expertise, and that would it would make more sense. I mean, you know, you've got, if you're partnering with Google, they've got deep mind and and you know these, these different arms of their company that are building powerful foundation models, and it's unlikely that that you're going to be able to build something to compete with them. And that's the logic behind building your own foundation model.

**Anna:** 50:54

Yeah. So where I would, where I think the distinction to draw is in when folks think about foundation models right now they're thinking about human language foundation models, and there is, you know. So what? What are the ingredients for a good foundation model? A lot of data, talent and a lot of compute, and so the data is largely out there, right? And there's common crawl. We can all get that data, and some folks have proprietary data, but that is largely been democratized in human language. Compute is a money problem, let's just assume. You know lots of folks can spend lots of money on compute, but some folks will decide to spend more, right? So that's one factor. And then there's talent, and I do think that that really matters In the biological domain. You can't take the data piece for granted in the same way that a lot of folks are taking it for granted, I think, right now in human language or where the money sort of dwarfs the data. Because the real question is do you have enough money to actually ingest as much data as is out there? And right now there's, candidly, just not not that much public data and biology relative to what you would really want. I do think that to your point, like we're very open to collaborating on this space, like I would not argue that on those three axes like compute, talent and data. Like we didn't invent transformers, right. Like we've got really, really terrific application scientists who are great practitioners of AI and will probably be a lot better at thinking through how this data needs to get ingested in things, because there's specialty knowledge on the kind of the ontology of biology, if you want to think about it that way. But we should absolutely be collaborating with leading thinkers, whether that's at DeepMind or Academic Labs or OpenAI or folks like that around what types of architectures are going to be game changing in this field, and I think we're very open to collaborating on building these models. Our view is just that you're going to need that data and even somebody like OpenAI right, like they had to partner to get a lot of that data with scale, and so Ginkgo does to some degree, play the role of scale AI in this ecosystem, and I think it does remain to be seen whether we're going to have to build our own foundation models or whether we can serve that kind of data, serve as that data partner role and collaborate with other folks on building models of that scale.

**Craig:** 53:20

Yeah, well, that's fascinating.

**Anna:** 53:23

It is just mind bending to think about the potential of biology, like if we actually just understood the stuff, and the reality is we don't. But the reality is also we're learning really, really, really fast, and so I think the pace of change is going to be pretty mind boggling, I think, and this is a very powerful substrate to be learning about, and so it's an honor, if nothing else, honestly, to be able to work in this space.

**Craig:** 53:50

That's it for today's episode. I want to thank Anna Marie for her time. If you want to read a transcript of today's conversation, you can find one on our website. I on AI, that's EYE-ONAI. In the meantime, remember the singularity may not be near, but AI is already changing our world, so pay attention.