[MUSIC]

**DAPHNE KOLLER:** I think one of the very, very thin silver linings around this very dire situation that we find ourselves in is that there is, I hope, a growing appreciation among the general public for what science is able to do for us today and how much of that ability rests on decades of basic science work by many, many people.

[MUSIC]

**KEVIN SCOTT:** Hi, everyone. Welcome to Behind the Tech. I'm your host, Kevin Scott, Chief Technology Officer for Microsoft.

In this podcast, we're going to get behind the tech. We'll talk with some of the people who have made our modern tech world possible and understand what motivated them to create what they did. So, join me to maybe learn a little bit about the history of computing and get a few behind-the-scenes insights into what's happening today. Stick around.

[MUSIC]

**CHRISTINA WARREN:** Hello. Welcome to *Behind the Tech*. I’m Christina Warren, senior cloud advocate at Microsoft.

**KEVIN SCOTT:** And I’m Kevin Scott.

**CHRISTINA WARREN:** Today, our guest is Daphne Koller, who’s the CEO and founder of insitro, which is a company that works at the convergence of biology and machine learning.

**KEVIN SCOTT:** Yeah, I’m guessing everyone has a newfound appreciation for how important biomedicine and biotechnology is at this time with the COVID-19 pandemic still raging around us. And I think Daphne is doing some of the most interesting work right now in the field that is, as we’ve seen with several of our other guests, like, this really powerful combination of biology and machine learning and high-performance computing and laboratory automation. Like, what they’re doing is really wonderful work. And Daphne is just sort of a brilliant computer scientist and has had many, many different chapters in her career that are inspirational.

**CHRISTINA WARREN:** That’s so true, Kevin. I cannot wait to hear this conversation.

**KEVIN SCOTT:** Yeah, so let’s get started.

[MUSIC]

**KEVIN SCOTT:** Our guest on the show today is Daphne Koller. Daphne is a machine learning pioneer. She’s CEO and founder of insitro — a company that applies machine learning to pharmaceutical development. Daphne was a computer science professor at Stanford, co-founder and co-CEO of Coursera, and is a MacArthur fellow.

She was also named one of Time Magazine’s 100 most influential people, and I’ve had a copy of her book — Probabilistic Graphical Models on my bookshelf for more years than I know. So, welcome to the show, Daphne.

**DAPHNE KOLLER:** Thank you, glad to be here. And I would be very impressed if you actually read the book.

**KEVIN SCOTT:** Well, I read the book many, many years ago when I was at Google, where your work was very influential in how we thought about doing some of our very early and what we thought at the time was very sophisticated machine learning work in the ad system. So, I was — I do have to admit, at the time that I picked the book up the first time, I was a compiler and programming language person, so it was not an easy read.

**DAPHNE KOLLER:** It is definitely a bit of a tome. At this point, it serves well as a doorstop if you need one. (Laughter.) It’s very, very big.

**KEVIN SCOTT:** So, in any case, I’m so delighted you could be with us here today. So, we typically start off these conversations with a bit about your background. So, I’m curious how it is that you got interested in science and technology in the first place.

**DAPHNE KOLLER:** So, I was interested in science ever since I was a kid. My family had this series of books that were the Time Life series on everything from asteroids to how plants grow, and I just used to sit and read them for fun.

I didn’t get interested in technology until my freshman year of high school, when my parents came here. My dad came on sabbatical to Stanford. And for the first time, I was in a school where there was a computer center. This was a long time ago so I’m going to date myself at this point, but these were TRS-80 computers that were time-shared across two people. And I got to learn to program. And I found it an amazing experience, where you could actually tell a computer what to do and it did it, which didn’t work for me in any other context. But here, it actually did.

And so, I got interested in computing at that point. And that I think led to basically my college choices and so on. And ultimately, I think combining the two in the work that I do now, but that’s many years later.

**KEVIN SCOTT:** So, on those TRS-80s, it sounds like your experience is very similar to mine, actually. Like, I — in high school, I did a bunch of coding on TRS-80s. So, was your language of choice there the BASIC interpreter, or were you using something else?

**DAPHNE KOLLER:** No, at that point, it was BASIC. This was the only language that was available when I was here in high school, but I very quickly migrated beyond that to Pascal and then C.

**KEVIN SCOTT:** Yeah, that’s really — that is almost exactly the same language — although, there was some assembly language in there because I wanted to code games and you had to supplement the BASIC with a little bit of assembly language if you wanted to make things move around on the screen.

So, you know, this thing that you said about being attracted to programming as a kid because the computer would listen to you, like, I think is very interesting. It is one of those things that I think can give kids agency. And, you know, I know that you — you know, both as an educator at Stanford and as one of the co-founders of Coursera, you’ve thought a lot about how to educate both kids and adults. Like, how important do you think that sense of agency is in getting kids interested in computing?

**DAPHNE KOLLER:** I think it’s very important. Kids, I think it’s difficult for us to appreciate as adults just how powerless most kids feel. Certainly, the ones from less-advantaged backgrounds, but even the others. And I think giving them an avenue where they can really dictate, if you will, what happens is super exciting for them.

And I think we are not giving them enough of that in how we currently teach technology because we’ve moved far away from the programming in how one teaches computers in most schools. And if you actually came back to that and say, “Hey, look what you can build,” and it actually works. I think it’s an incredible feeling of empowerment for kids.

**KEVIN SCOTT:** Yeah, one of the things that I’ve struggled with, with my own kids, is trying to get them interested in programming. So, I’m not trying to force them to learn anything that they don’t want to. Like, we try to expose them to a bunch of things. And it took a while to figure out how to find an entry point into coding that was interesting for them.

And the thing for my son was Roblox. So, which is this game that he plays on his tablet obsessively. And as soon as he figured out that there was a way for him to create his own stages in Roblox, like, that was the thing that enticed him to want to program.

**DAPHNE KOLLER:** I think it’s become harder to get kids interested in programming because the programs that are already out there are really sophisticated and fancy. And what kids can create is always going to pale by comparison to what is already out there, which is not a problem that you and I had when we were starting with computers. You couldn’t — you didn’t have a tablet on which you could play amazing games.

And so, what we created seemed kind of cool. And now when kids create, it seems not quite as cool as the games they have on their phone. And so, the question is: Is there a way in which we can give kids that same sense of excitement about what they’re creating so that it does seem cool and interesting? And I don’t think we’ve paid enough attention to that.

**KEVIN SCOTT:** Yeah, and you know, it’s interesting that you bring that up, because we have talked with a bunch of other guests on this podcast where it’s also true that the programming tools that you have available to you now are vastly more powerful than the ones that we had when we were first learning to program. But it could very well be — and like 100% agree with what you’re saying, that the gap between the sophistication of the software has maybe grown even further apart from the power of the tools, you know, for entry level — for kids.

**DAPHNE KOLLER:** Yeah, no, that’s right. And my daughter, you know, she was kind of interested in machine learning for a while, but — so I said, “Well, why don’t you try your hand at one of those Kaggle competitions?” And the problem is that the Kaggle competitions, they’re full of really sophisticated, top-notch programmers looking to build a reputation so that they can go get jobs because of their machine learning street cred. And kids like my daughter have no chance of even getting into the top 100.

So, it was kind of a bit of a demoralizing experience in the sense of nothing she could do would ever rank up there. And so, I’m not sure it ended up serving a positive purpose. And so I wonder if it would make sense to have a Kaggle for kids or something that would let kids compete in a playing field that was more even and would get them excited about going on to the next level.

**KEVIN SCOTT:** Yeah, I think that’s a fantastic idea. I mean, and we understand how to do this with sports, right? So, like, there are all these sports leagues and sport opportunities for kids where you can get them into a team where they can learn the sport and get all of this physical activity, but they’re not, like, completely and utterly outmatched either with their teammates or the teams that they’re playing. So, like, it just seems perfectly reasonable to me that we could figure out how to do this with some of these coding competitions.

**DAPHNE KOLLER:** You would think.

**KEVIN SCOTT:** I mean, if you weren’t already, like, busy founding, you know, running a company, I’d say that sounds like a good thing to, you know, to go do. (Laughter.)

**DAPHNE KOLLER:** This will be my second or third alter ego if I were able to wrangle that, yeah.

**KEVIN SCOTT:** Yeah, indeed. So, you learned to program when you were in high school. And when you went to college, did you choose computer science?

**DAPHNE KOLLER:** So, I actually had an interesting early career in that regard because I actually was young when I was doing high school and I then started — when I came back from the United States after that sabbatical, I actually started college in parallel with high school because I’d always found high school to be not as inspiring as I would have hoped, and I found a lot more flexibility in college curriculum.

So, I started to study math and computer science while I was still finishing up high school. And yeah, I think it was the timeliness of that in that I had just come back from the United States, where I’d learned to program, probably influenced my career choice. And who knows if I’d waited three or four or five years, like most kids do, then maybe I would have picked something else.

But the nice thing that I found about computer science and even more so over time is that it’s actually an entry point into multiple other fields, because especially today, but even back then, most fields can benefit from computational methods and using computer science. Whether it’s algorithmic thinking or now even more so machine learning, all of them find that technology a really useful and often very divergent way of approaching the field.

And so that allowed me in my career to touch on so many different areas from things that are more core tech like robots and computer vision, to things that are a little bit more distal, but at least today, still considered more core, like natural language processing, to things that are a little bit less viewed as part of core computer science.

I did a lot of game theory, for instance, and economics early in my career. And now I’m doing a ton of science and medicine. And it’s not that I’ve become a biologist. I am still a computer scientist, but the tools are just so useful in all these different disciplines that as a computer scientist, I’m not only able to do interesting biology, I’m able to do it in a way that is often very different to how someone who was trained as a biologist would approach that same problem.

**KEVIN SCOTT:** Yeah, so what was the first interesting thing either as a high school student or when you were in college that you did that wasn’t you know the core CS stuff, like operating systems, compilers, algorithms, data structures — where you sort of realized, like, “Oh, wow, like, this computer science stuff that I’ve learned is, like, a super power that lets me do a whole bunch of things?”

**DAPHNE KOLLER:** So, I think the first one was really the integration between game theory and computer science and the meeting point in distributed systems. So, that was actually what my master’s thesis was about was really trying to understand how one could view a distributed system as a multi-agent system, where the agents had their own motivations, their own goals and utility functions.

And that started out towards the end of my master’s degree, and then came back when I started my PhD, where I was, at that point, trying to explore it from both sides. Can an understanding of the multi-agent incentive function, if you will, from the game theory perspective, help us build better distributed systems? But also, conversely, if game theory was a really interesting framework for decision-making, can we use the tools — the algorithmic tools that computer science gave us to help us find better solutions to game theoretic problems that were not even necessarily within the scope of computer science? So, can we help people make better decisions in the multi-agent setting by turning it into a computational problem so it wouldn’t be this kind of bespoke, somewhat obscure mathematical analysis that only game theorists could do, but actually a tool that’s useful in decision-making?

And so, that was kind of the entry point for me that went actually from decision-making and multi-agent systems, to decision-making in single-agent systems, to modeling of the world that would enable decision-making, to then learning those models from data, which is what took me to machine learning. So, that was actually that trajectory.

**KEVIN SCOTT:** And did you have anyone or, like, any interesting waythat you were getting into these tangential fields? So, like game theory, for instance, like, was this something where you had an influential mentor who put you onto it? Was this you just independently getting curious and reading a whole bunch of stuff? Like, what’s your approach to, like, learning these disparate things?

**DAPHNE KOLLER:** I wish I could tell you that it was a systematic, thorough exploration where I took a broad perspective and tried to figure out what was interesting and most useful. Often, it’s a bit of serendipity and just affinity to a particular space and maybe just a sense of --there’s something here that could be exciting. On the game theory side, I just happened to take as part of my undergraduate degree, which was a dual degree in math and computer science, there was a game theory class. And so, I took that and I got really intrigued by the truly elegant mathematics that underly it. And I said, “Wow, this could be a really cool way of thinking about interactions in a computer system.”

So, that was purely serendipity. My move into biology was much later. Interestingly, my father was a biologist, and so I had always actually steered away from biology, partly I think because like most kids, you don’t want to do what your parents do. But also because at that time, when I did take a biology class, it was incredibly descriptive. It was like a catalogue of, you know, obscure Latin names of plants or pieces of cells, and it was all about memorizing that this does this to that. And I was just completely uninterested in doing that because it seemed like there were very few principles in play. It was all about the details, and I’m not good at details. So — especially memorizing them.

So, I didn’t really get into that and did my entire both high school and undergraduate career focusing on things that were much more interpretable in terms of principles and systems. So, math and physics and computer science.

And the reason I got into biology and medicine was actually when I came back to Stanford as a faculty member and started to do machine learning and realized just how boring and uninspiring the data sets that we machine learning people had to work with at the time that machine learning was getting off the ground. I mean, one of the flagship data sets, quote/unquote, was something called the 20 News Groups, which is exactly what it sounds like — it’s articles from 20 very boring news groups, and you have to classify which news article came from which group. And it was not interesting technically and it certainly wasn’t very aspirational.

And so I started to look around, what other interesting data sets were around, and specifically, my focus at the time was on data sets that were more richly structured — relational data sets, if you will, where there’s multiple types of entities, multiple types of relationships, and looking for data sets that had those characteristics that were available. And most of those were trapped behind the doors of companies that weren’t very excited about making those available to outside researchers.

And biology at that point was, like, “Oh, well, look there’s genes and cells and proteins and people.” And I actually started early work on, interestingly enough today, on epidemiology, at that point, of tuberculosis and tracking infection chains and figuring out if you could sort of pinpoint where an infection started — something that seems very timely today. But at that point, the data sets there were also pretty small, but they were much more interesting than the 20 news groups.

And so, I started to work first on things like the TB epidemiology and then on some of the earliest data sets that measured the expression or activity level of different genes and different types of cells, and that, of course, was a network problem galore, because you really had to figure out that this gene did this thing to this other gene, and so that really created a much more interesting technology challenge.

And then from there, I actually started to get interested in the biology in its own right, because it was not only more interesting, but it was also much more aspirational in terms of what you could do with it actually could help people. So, and then that grew to be a much more significant driving force for me over time, the wish to just do good, not just good science, but also good to the world.

**KEVIN SCOTT:** Right. And so I want to dive deep in just a minute into this — both the biology and in this notion of, like, how it is that we technologists can be doing more to do more good in the world. I want to take a moment and double-click on this point that you just made, which is this very strong correlation between what people will do research on and what people will study in machine learning and the available data, because I know one of your former colleagues, Fei-Fei Li, like, helped put together ImageNet, which catalyzed a whole bunch of, like, really interesting developments in computer vision.

And it’s just true, like, the data sets that you have available to play around with will sort of dictate the character of your research. So like, in a way, you were doing something extraordinary at the time by realizing, all right, well, I’m going to go find the interesting data. And I do think as part of this notion of we should be directing our efforts towards things that will do public good is partially about making sure that people have the compute resources and tools and whatnot, but like, it’s also about making that data available — data sets that are relevant to the problems we want solved.

**DAPHNE KOLLER:** Absolutely. I mean, when you think about an aspiring young researcher or even an undergrad or even one of those, like, high school students that we talked about earlier, having a data set that is interesting, that offers potential for them to do something innovative and cool and that is processed and easily accessible and where the — kind of — there’s at least initially a set of well-defined problems that they can tackle while they explore the data to come up with potentially new problems, that I think is such an important entry point for people into the field.

And we all understand that that’s not where the real action is, ultimately. If you’re going to become a leading researcher in the field, part of what you need to do is really develop your own way of finding data sets. Although, to be fair, there are some incredibly talented people who just continue methods development on data sets that other people have already created. And I think that’s a very worthwhile path as well.

But, so, for either of those, giving people and easily accessible, first entry point into a field is just absolutely critical, as opposed to what I’ve seen in a lot of early-stage machine learning projects, where they tell people, “Oh, go around, figure out what problem you think would be interesting for you to solve, and then figure out how to get data for it, and then figure out what machine learning algorithm is good for it.” I mean, that is such an insurmountable mound of stuff for someone to tackle the first time they’re getting into the field, that it explains why we see one of two things — people moving away from the field and not doing that, and that’s especially true for people who are not quite as privileged as others in terms of what we give them as a starting point.

Or they go work on the same old data sets as everyone else, which you know, I think Fei-Fei’s work on ImageNet was actually transformative. But at this point, I’d like people to start thinking about other forms of data that they get — that they could get practice on, and we haven’t — there haven’t been enough Fei-Feis to go and create those data sets in other places.

**KEVIN SCOTT:** Well, and you can sort of see it even in how we reward folks. You know, I — like maybe this is sort of a controversial thing to say, but like I was a little bit shocked that Fei-Fei wasn’t on the same roster of folks who got the Turing award for deep learning because the ImageNet stuff was like potentially — I mean, it was absolutely a precondition for, you know, the stuff that, you know, Hinton and Lacune and Benjio did.

You know, whether or not folks actually agree with that is almost beside the point that we don’t recognize this data collection and, like, building these data assets as much as we do the fancy algorithms.

**DAPHNE KOLLER:** I think there has always been a lot of appreciation in the machine learning community as a whole for technical fire power, for yet another improvement on algorithms or models that, admittedly, is an amazing contribution. And, obviously, a lot of those developments have been what’s opened the door to the performance that we see today. But there’s been less appreciation I think for the intellectual endeavor of doing work that is more applied.

And I think people often don’t understand the amount of intellectual endeavor and thought that goes into questions such as, “What is the right problem within this big sea of a space, like biology or earth science or whatever? What are the questions that are both technically tractable and yet can be transformative to what the field is trying to accomplish?” And that is an incredible intellectual exercise, followed by the second intellectual exercise, “Well, if this is the problem that we aim to solve, how do we get the data to actually solve it? And can we acquire it? Can we clean it? Does these — do these data sets have issues that we need to address, or do we need to go and collect data *de novo*?”

That, too, is an incredible intellectual exercise and often a very time-consuming feat. And I agree with you that those efforts are not always as recognized as some of the sort of mathematical or machine learning sort of flashier efforts.

**KEVIN SCOTT:** Yeah, so — which is a really good segue into what you’re doing right now, which is applying machine learning to a very, very worthy set of problems. And I’m guessing you started well before we were in this pandemic moment that we’re in right now. But what you’re doing, I’m guessing, is more relevant now than it was even six months ago.

**DAPHNE KOLLER:** No, absolutely. I think one of the very, very thin silver linings around this very dire situation that we find ourselves in is that there is, I hope, a growing appreciation among the general public for what science is able to do for us today and how much of that ability rests on decades of basic science work by many, many people that— much of which is publicly funded work at academic institutions. Without that level of progress that we’ve made, the concept of, say, creating a vaccine in 12 months would have been completely ludicrous a few years ago. Or the work that’s being done on repurposing of drugs that exists to help address some of the — even if not cure the disease, at least slow its progression or help ameliorate some of the more significant inflammatory consequences.

There is thousands of drugs out there. You can’t do thousands of clinical trials for each of them, so a lot of the work that we’ve done on interpreting cell-based assays and understanding the immune system and understanding things like cytokine storms and such, those are all key building blocks for the fact that we actually have at this point two drugs, and hopefully more coming, that at least are somewhat helpful in addressing this disease. And so, I’m really hoping that people are paying attention, that science matters. It really matters. And you should be supporting science and listening to science in the good days, because when the bad days come, it’s going to be too late to sort of suddenly realize that you need science. So, sorry. That was my little soapbox right there. But —

**KEVIN SCOTT:** I think it’s — everything you said, I could not more strongly agree with. And, you know, like, one of the things that I’m hoping for, like, this is my desired silver lining potentially for this moment that we’re in is, like, I think we are making very rapid progress towards vaccines and therapeutics and better understanding exactly the mechanism of this miserable little virus.

But I’m hoping, like, we will, in this moment, see how much science can accomplish when we point it at a task like this and, hopefully, we will decide that that is a worthy set of things to invest much more in than we have been over the past couple of decades.

**DAPHNE KOLLER:** No, I completely agree, because I think with a concerted effort like what we’re currently putting into the coronavirus and the remarkable advancements that have happened in both science and technology over the past, I would say decade, because the last decade has been transformative in science in many of the same ways that has been transformative in machine learning, but coming in it — into it from the other side. I think we have a chance of making significant headway against other diseases that are currently still scourges that are incredibly damaging and shorten people’s lives, reduce their quality of life. And I think we, with the right investment and the right focus, we could actually make a difference.

**KEVIN SCOTT:** So, tell me a little bit about what you’re doing at insitro.

**DAPHNE KOLLER:** So, the premise for what we’re doing really emerges from what I said a moment ago, which is that this last decade has been transformative in parallel on two fields that very rarely talk to each other.

We’ve already talked about the advancement on the machine learning side and the ability to build incredibly high accuracy, predictive models in a slew of different problem domains if you have enough quality data.

On the other side, the biologists and bioengineers have developed a set of tools over the last decade or so, that each of which have been transformative in their own rights, but together, they create, I think, a perfect storm of large data creation — enabling large data creation on the biology side, which when you feed it into the machine learning piece, can all of a sudden give rise to unique insights.

And so, some of those tools are actually pretty special and incredible, honestly. So, one of those is what we call “induced pluripotent stem cells,” which is — “we” being the community, not “we” at insitro — which is the ability to take skin cells or blood cells from any one of us and then by some almost magic, revert them to the state that they’re in when you’re an embryo, in which they can turn into any lineage of your body.

So, you can take a skin cell from us, revert it to stem cell status, and then make a Daphne neuron. And that’s amazing, because that Daphne neuron carries my genetics. And if there are diseases that manifest in neuronal — in a neuronal tissue, you will be able to potentially examine — assay those cells and say, “Oh, wait, this is what makes a healthy neuron different from one that carries a larger genetic burden of disease.” And so that’s one tool that has arisen.

A different one that also is remarkable is the whole CRISPR revolution and the ability to modify the genetics of those cells so that you could actually create fake disease — not fake disease, because it’s real disease, but introduce it into a cell to see what a really high penetrant mutation looks like in a cell. And then, commensurate with that, there’s been the ability to measure cells in many, many, many different ways, where you can collect hundreds of thousands of measurements from each of those cells so you can really get a broad perspective on what those cells look like, rather than coming in with, “I know I need to measure this one thing.” And you can do this all at an incredible scale.

So, on the one side, you have all of this capability for data production, and on the other side, you have all this capability for data interpretation. And I think those two threads are converging into a field that I’m calling “digital biology,” where we suddenly have the ability to measure biology quantitatively at an unprecedented scale, interpret what we see, and then take that back and write biology, whether it’s using CRISPR or some other intervention to make the biological system do something other than what it would normally have done.

So, that to me is a field that’s emerging and will have repercussions that span from, you know, environmental science, biofuel, bacterial or algae that do all sorts of funky things like suck carbon dioxide out of the environment, better crops, but also, importantly for what we do, better human health.

And so, I think we’re part of this wave that’s starting to emerge. And what we do is take this convergence and point it in the direction of making better drugs that can potentially actually be disease modifying, rather than as in many existing drugs, just often just make people feel better, but don’t really change the course of their disease.

**KEVIN SCOTT:** And so, this technology that you’re talking about, will it be used to make the drugs or to examine the effect of potential drugs or both?

**DAPHNE KOLLER:** Both. So, it actually starts with understanding where you even want to develop drugs for. So, a lot of the problems that we have with current-day drug failures, which are, depending on which statistic you believe, the success rate of the drug discovery effort from beginning to end is somewhere around 5%. So, think about that, it’s a 95% failure rate.

And a lot of this is because we just don’t understand the biology, we don’t know where to develop drugs towards. What is the right target and what is the right cell type and what is the right patient population? So, it starts with predicting, using machine learning, what viable targets are in the context of a given disease and a given target population. And then from there, okay, how do we design drugs more rapidly so that we don’t have to wait, you know, five years for — or sometimes much longer for a drug to emerge. And so really want to close that arc of going all the way from the biology to the actual drug.

**KEVIN SCOTT:** There’s so much obvious potential for this thing that you’re calling digital biology and, like, there are a bunch of very promising companies and a bunch of, like, very brilliant researchers who are doing work in this area. So, I’m curious if you have any thoughts on what are the obstacles standing in our way of going faster? Is it educating the right people? Is it we need more data, we need more compute resources, we need breakthroughs in particular areas? So, like, how do we make all of this go faster?

**DAPHNE KOLLER:** So, I think “yes” to everything that you said, with the possible exception of more compute power. I don’t think that’s currently the rate-limiting aspect.

**KEVIN SCOTT:** That’s great. You are then an unusual part of machine learning. (Laughter.)

**DAPHNE KOLLER:** Well, I mean, maybe I’m being overly optimistic, but there is just you know, you can currently turn on the tap and pay your cloud provider, whoever that is, more money, but it’s not like that’s the place that is currently blocking us. What’s currently blocking, I think, is — working my way backwards through your list — is having not only more data but having the right data — data that really helps inform the answers to the questions that are really going to transform the space.

Creation of biological data is challenging. Those are living beings that you’re manipulating, and as such, there is all sorts of funky things that can go wrong that those of us who were trained as engineers with man-constructed artifacts are not familiar with that, you know, cells behave differently for reasons that we do not understand. They clump. They get infected with these things called mycoplasm that ruin your whole experiment and infect other cells.

There’s just so much stuff that can go wrong in a biological experiment where you manipulate living beings that you need to be really good at it and you need to be very, very careful in how you do the experiment, but equally careful in figuring out what experiment is it that you want to do. Because experiments take time, there’s only so far you can accelerate a cell in getting it to grow faster. And even more so when you’re dealing with a larger living organism, be it a model system or a human. So, the experiments are much more high stakes because it’s not just a matter of, okay, let’s push a button and launch another 10,000 of those in the cloud.

And then I think working our way backwards, in order to really answer those questions in the right way, which is what are the experiments that we need to perform? The ones that are going to be truly meaningful, transformative, feasible from an experimental perspective, and at the same time feeding into the machine learning in the right way, you need to have at least a group that speaks both languages, that understands the biology in terms of what’s useful and also what’s possible, and at the same time a group of people on the computer science side who understand what the technology can do and where to find within that sort of stew of the broader field of, say, biology or even drug discovery, problems that are both impactful and tractable.

And those people who speak both languages are very few and far between. There is maybe a few more of them coming up as educational institutions become more cognizant of the need to train interdisciplinary people, but those people are very hard to find. They — if you talk to your average computer science person or machine learning engineer and you put them in a room with your average biologist or medical doctor, they could — and even if they come in with all of the good intentions of wanting to collaborate, they have not only completely different languages, they have completely different mindsets.

So, coming back to some of the earlier points that we made, biology still even today is a lot about the details. And the reason for this is that the exceptions, those little nit-picky things that don’t line up with everything else that you’ve seen are often the starting point for new discovery. So, people kind of want to look for those, whereas engineers really care about let’s find the principles that cover 95% of what we see, because that’s going to be good enough for us to go and build systems.

And so that mindset or those two mindsets are so at odds with each other in many ways that getting people to really communicate in a way that is collaborative and constructive is really hard. And if I can point to the one thing that we’ve done at insitro that I’m the proudest of is that we’ve built a community of people that span a broad spectrum of disciplines in that range and are actually working as a single team. And that’s just very unusual.

**KEVIN SCOTT:** Yeah, that’s — I mean, fascinating. I’m just sort of curious, like, what is — you know, and this may, you know, require going out on a limb you don’t want to go out on. But you know, one of the things that’s made computing so much more powerful over the past five decades, like the entire course of modern computing history is that we have this way of building abstractions that compose, where we don’t have to understand all of the little nit-picky things. I mean, it’s useful to have a model for the nit-picky things when your abstractions fail so that you can go investigate things and figure out what went wrong.

But, you know, by and large, you’re sort of trusting a bunch of very powerful, very high-level abstractions when you go do your job as a computer scientist or a software engineer today. You know, everything from, like, I can just sort of push a button and a virtual machine materializes on a server or in the data center somewhere, in the cloud and like I don’t have to, you know, worry about all of the just colossal amount of complexity that makes that happen. Is there an equivalent mechanism at play in modern biology?

**DAPHNE KOLLER:** You know, it’s interesting that you bring that up because of-- maybe not that surprising because we were both trained as computer scientists, but one of the things that I love about modern biology is that we’re getting there. So, there’s an emerging set of building blocks that are relatively well defined in terms of what I’m going to call their API, which is obviously not a word a biologist would ever use, but they have a well-defined kind of input/output functionality.

And these include things like CRISPR for genome editing, where you can basically say, “Okay, this is what I want to do to edit the cell, and then I do that, and there’s a set of steps that we need to do, and then an edited cell comes out.” So, that’s the “glass half full” side of it, that there are these building blocks that are emerging and you can start to compose them and do more interesting things with larger and larger, more complex programs, if you will, that are written in terms of those building blocks.

The bad news is that each of those building blocks is, in turn, based on a system that is not a nice, predictable, well-understood system like a computer. It’s something that involves living cells, and so everyone I think has heard about the risk of, say — I’m taking a very simple example, off-target effects of CRISPR editing.

And the fact that — which off-target effects you get depends on many things that we don’t understand. Not only which cell type it is, but the specific individual from whom it came, gives rise to somewhat different consequences. The state that the cell was in at the time that the experiment was started, so you can think of these as on the one hand composable building blocks that you can start to sort of create systems with. But each of them is incredibly variable in its response, so it creates a distribution of outcomes that we really don’t understand. And we need to design these experiments in a way that is robust enough that it’s hopefully useful, even despite that variability, and put in what we as computer scientists would call QA pieces that measure as many of the pieces along the way that we possibly can in order to figure out what emerged from each of those building blocks so that we can trace the repercussions down the line. And it’s very hard.

So when you ask what is it that makes this hard, it’s that you have to bring that systems mindset of QA and tracking and putting in incredibly stringent sort of constraints on each of those building blocks in the same way that you do when you build an Intel microchip fab, for instance, to a discipline that really hasn’t done as much of that, but in a way that is cognizant of all of the sources of variability and errors that might occur in a biological system. So that confluence is really hard to put together.

**KEVIN SCOTT:** Well, it strikes me that this bag of techniques that you are bringing from your background — so probabilistic modeling and machine learning they are the best possible contemporary set of techniques for dealing with some of these uncertainties. Whereas if you had to go in and, like, describe these systems with a set of partial differential equations, you’d be lost from the outset.

**DAPHNE KOLLER:** I completely agree. I mean, unfortunately, our ability to describe biological systems using rigid mathematical — deterministic mathematical tools — fails once you go beyond the atomic level, and even there — I mean, when you think about something that is relatively circumscribed, like a single protein folding, you can do some of the differential equation modeling, but even there, we’ve seen techniques that take a step back and say, “You know what? Let me not try and construct detailed, mechanistic models, but instead, let’s give the machine enough data to learn from and it’ll pick up patterns that might be useful.” That’s what made the deep fold success from DeepMind work is that they took a machine learning approach to that.

Now, the critical piece, of course, and that comes back to our conversation a moment ago is that they had enough data to train on folded proteins. And getting enough high-quality data is what it’s all about in this new world of bringing machine learning into this space. And that’s why we built insitro the way we did.

**KEVIN SCOTT:** Awesome. Well, we’re just about out of time, and I wanted to ask you before we wrapped up, what do you do for fun? So, you have, like, what sounds to me like an incredibly fun job, but like, there must be something outside of work.

**DAPHNE KOLLER:** Well, so, first of all, I am grateful to have a job that is as much fun as this in the sense that I get to read all of the coolest papers in biology and all the coolest papers in automation and in machine learning and figure out how to put them together in new ways and do it towards a goal that I think is just truly important, which is how do we make people healthier? Which, to me, is — and I’m going to go on a soapbox for just a moment and talk about the fact that I think part of our goal here as we you know — when we were put on this earth was to try and leave the world a little bit better than it was when we came into it.

And we should be doing that. And for those of us who had the privilege of being born to relatively affluent, well-educated families, where we didn’t need to struggle for where our next meal is coming from, that burden is actually even higher and we — we should be thinking about how we can give back. So, anyway, sorry, that was —

**KEVIN SCOTT:** No, that’s so important.

**DAPHNE KOLLER:** Yeah, so, but that being said, the thing that I most liked to do for fun pre-coronavirus was to travel and see parts of the world that are different from the little cocoon where we live.

So, I’ve been to 65 different countries so far, six different continents. Have not yet been to Antarctica, that’s definitely on the bucket list. And I find it to be a wonderful experience both in visiting other cultures and seeing how different people live. But also, I love being out in nature and the outdoors and hiking and SCUBA diving and sailing and doing all that.

So that is the thing I used to do for fun. I have no idea when the next time I’ll be able to do that is, unfortunately, at this point in time. So, the other things that I like to do is just spending time with my family and, you know, going on local hikes in nature, which are not, perhaps, as dramatic as visiting Iceland or the Great Barrier Reef or this incredible lake in Palau that has jellyfish that don’t sting and you can swim in them.

**KEVIN SCOTT:** Wow.

**DAPHNE KOLLER:** But at least it’s being outdoors in the fresh air and I’m lucky enough to live in a part of the world that has some beautiful scenery even locally, so I go for hikes a lot these days.

**KEVIN SCOTT:** Yeah, well, hiking in northern California is not bad at all.

**DAPHNE KOLLER:** Nope, can’t complain too much relative to what the situation could be. But I do wish we could get back on a plane at some point and visit some of those amazing places elsewhere in the world.

**KEVIN SCOTT:** Well, I’m hoping that probably not as soon as we want, but sooner than we would ever have been able to do at any other point in human history, science will be able to give us enough safety around coronavirus that hopefully you’ll be able to travel — travel soon. I won’t make any predictions about when “soon” is, but like, let’s hope for soon.

**DAPHNE KOLLER:** Very much hope so. And I think if we do get to that point in the near term — and by near, I mean within the next 12 to 18 months — I hope people will appreciate the miracle that it is and the many decades of work by so many people that needed to happen in order to make that possible.

**KEVIN SCOTT:** Yeah, and I think that is the perfect place to stop. So, thank you so much for being on the show today. This was a fascinating, fun conversation and I’m glad we got to talk to you today.

**DAPHNE KOLLER:** So am I. Thank you very much.

**KEVIN SCOTT:** Awesome.

 [MUSIC]

**CHRISTINA WARREN:** So, that was Kevin’s conversation with Daphne Koller, CEO and founder of insitro. And, oh, my gosh, that was so interesting. There were so many amazing parts of that conversation. I — I’m not even honestly someone who’s that into biology, and there are so many things that I’m going to think more about and that I want to kind of pull on more strings based on that conversation. That was amazing.

**KEVIN SCOTT:** Yeah, I think one of the really great things about Daphne and one of the things that has made her such a great scientist and entrepreneur is that she thinks about everything that she does extremely deeply. She has this wide-ranging curiosity, which I think is — you know, one of the best superpowers. Like, you combine that with persistence and you find yourself in all of these situations where you are making connections across disciplines and doing a whole bunch of things that maybe you wouldn’t be able to imagine if you were a more narrowly focused, person or had a more narrowly focused set of interests.

And, like, she said so many things in that conversation that I’m, like, “Wow, I really need to go think about this more deeply myself.” Like,— just one of the, casual things that she said was this need for, like, maybe a — you know, an equivalent of little league sports or, like, a kids’ Kaggle competition so that you can find the right competitive and social dynamic for kids, getting themselves onboarded into machine learning. It’s a great idea, like —

**CHRISTINA WARREN:** Yeah, it really is.

**KEVIN SCOTT:** --somebody needs to go do that now.

**CHRISTINA WARREN:** No, I’m in total agreement, yeah, because, you know, we have little league and we have, you know, other sorts of competitions and when — when kids get older, there are some more science type of competitions, but to have something gamifying things when you’re younger around machine learning would be brilliant. That’s a brilliant idea.

And I love, you know, kind of her origin story, you know, the fact that she was writing her thesis on, you know, game theory and distributed systems and multi-agent incentive systems. Like, I was just like this — this is brilliant. You know, these are things that you — that to your point, you would need certain curiosity and just wide-ranging interests and persistence to really want to pursue.

One of the big take-aways I kind of got from this was something she said, you know, about how much science matters. And what are your thoughts about that, especially in the moment that we’re living in right now?

**KEVIN SCOTT:** Well, I think the thing that she tried to draw our attention to several times is that we — for this pandemic, and I think in general, like, we’re more dependent now on science to solve some of the really big problems that we are facing as a society or some of the challenges that we have to overcome in order to, you know, live our best lives and to have the future that we all want.

And the thing to remember is none of this is sort of overnight. Like, science is just years of substantial investment in a wide variety of things that build this foundation that when you get to a moment like the one that we have right now, you have all of the things that you need to go tackle these problems. So, if you don’t do these long-term investments in these foundational pieces in educating scientists and, like, giving them the ability to go do this work that builds this solid, solid foundation and, like, carries the whole field forward, you really can get yourself into a situation where a crisis comes along and, like, you just don’t have any way for science to help solve it.

And so like, I think that’s the thing that we all really need to remember, you know, when and I think it’s a “when,” not an “if,” when science pulls our butts out of the fire on COVID-19, like, we really need to remember that that is because we have for decades invested substantially in science, and hopefully it will redouble our resolve to go make even bigger investments in those foundations for the future.

**CHRISTINA WARREN:** No. I think you’re 100% correct. We need to continue to make these investments and I love that there are people like Daphne who are taking these two different fields, you know, taking computer science and machine learning, as well as biology, and working together so that hopefully the investments that are taking place now will be so beneficial decades to come.

**KEVIN SCOTT:** So ~~I —~~ I really do feel like we who are participating in the creation of science and technology have a real responsibility to society to focus on the right problems, for doing things that will produce positive benefit for all of humanity. And, you know, we just were — to use Daphne’s words, I think we were put here to try to leave the world a little bit better than we found it.

**CHRISTINA WARREN:** Absolutely. Absolutely. All right, well, that’s all for us today. Thank you again to Daphne Koller, and we are so glad that you joined us. We’ve learned so much information, and we hope that all of you at home got a little bit of nugget to impress all of your friends at your next socially distanced gathering. I know I definitely did. I’m definitely going to be dropping things like “digital biology” in conversation now. And remember to reach out to us anytime at BehindTheTech@Microsoft.com. Stay safe and be well.

**KEVIN SCOTT:** See you next time.