CRAIG: Hi, I'm Craig Smith and this is Eye On AI.

Before we begin, I want to propose to those listeners who are informed but not experts in AI, a quick way to get up to speed. The people at the application design and development company Infinite Red, who are supporters of this podcast, are offering a free mini-course, AI Demystified. It's a great place to start. At the end of five days, anyone will be able to speak intelligently about supervised, unsupervised and reinforcement learning and about practical business applications for AI. If you're a business leader or have friends or family that you would like to understand what you're talking about at the dinner table each night, recommended to them. I think everyone who wants to participate in the global conversation about our future should know the basics of AI. Check it out at www.learn.infinite.red.

CRAIG: Machine learning is transforming many industries, but none with such direct impact on humans as the healthcare space. One of the ways the powerful technology is being applied is in drug discovery by helping to identify compounds to treat disease. Daphne Koller, former machine learning faculty at Stanford and co-founder of the online education company, Coursera, has turned her attention to the field. She founded Insitro a company using machine learning to develop new drugs. We talked about her approach in identifying the cellular or genetic targets for treatment and some of the other ways that machine learning is being applied to drug discovery. As she says, the field is just getting started, but it promises to speed the development of new and better therapies to treat disease. I hope you find the conversation as interesting as I did.

CRAIG: Who is Daphne Kohler?

DAPHNE: I was an academic at Stanford. I was one of the, I think the earliest machine-learning hire into the department. I thought I would retire as an academic and then over time I became increasingly interested not just in theoretical aspects of machine learning, but actually in applying it to things that mattered to real people and started to work on applications to biology, applications to medicine, and over time that took me out of Stanford to do something that had more of a real-world impact.

DAPHNE: I transitioned from Stanford to actually go found Coursera in 2012 after we launched at Stanford the three big massive open online courses. I ended up building and leading Coursera for five years. It's still, I think one of the world's, maybe the world's largest online education platform in the higher ed space, and then after about five years, decided that I wanted to come and do something that was part of the machine learning revolution, but applied not to eCommerce or advertising, but to something that really made a difference and so came back to applying machine learning in the healthcare space.

CRAIG: Your education then was at Stanford before you became faculty there or do you study elsewhere?

DAPHNE: No, I did my PhD at Stanford. I graduated in '93- '94. I did a postdoc at Berkeley for a couple of years and then came back to the Stanford faculty in 1995. Half of my work was in core machine learning techniques and the other half was in machine learning applied to biomedical datasets. I worked on a variety of things that included active learning. How does a computer ask questions about examples that it's confused about so that it could learn faster. I worked on learning with really complex datasets that had a lot of structure, things that weren't just labeling the instance with one label like cat or dog, but something that had a much more richly structured label. That actually ultimately was the reason that made me interested in biology, because biology is such a richly structured domain and you're not just usually interested in just categorical label for the instances. You're looking for something that is much more of an understanding of the biological processes in play.

CRAIG: That's interesting. So this is all supervised learning primarily?

DAPHNE: No, I worked in both supervised and unsupervised and semi-supervised learning over my time at Stanford. In fact, one of my core research areas was in the space known as probabilistic graphical models where a lot of the work goes not in labeling a particular target variable, but rather in extracting from the data some kinds of representation of the interactions structure between key variables and the domain to gain something that both makes good predictions, but also gives you insight into the structure of what's going on. So for instance, that gave rise to a lot of the early work that I did on reconstruction of biological pathways for instance, from gene expression data. So I was working with a number of different types of data. So on the biology side, I was working with, for instance, activity levels of multiple genes across multiple patients. The goal was to try and understand how they work together and what was driving the changes in gene expression.

DAPHNE: So how does one gene potentially a regulator increase the level of a group of genes or decrease the activity level of a group of other genes? How does that interact with for instance, individual genetic variation across different humans? So what variants in the genome might be causing people to respond differently? So that's a very richly structured problem and there is no in many cases, in most cases I would say ground truth labels that one could learn from. So the question was here's a whole bunch of measurements that we have. How does one create a model that really explains the data in the best possible way? That particular line of work would fall into the category of unsupervised learning in the sense that there are no labels and the goal is simply to extract a structural model that predicts well the patterns that we see in the data.

DAPHNE: So the simplest example of an unsupervised learning problem, which is not what we did there, but just to illustrate, is clustering where you have a bunch of data points in some space and you're looking at it and you say, Oh, there is a group there that looks like they're kind of similar to each other and another group somewhere else that looks like they're kind of similar to each other and there is no labels. But by looking at it, one discovers patterns that are then useful for providing both insight as well as potentially additional predictions down the line. So what I described about the biological networks really fits much more into the category of unsupervised learning.

CRAIG: I see. And then how did you start in Insitro?

DAPHNE: So after I left Coursera in 2016 I realized that I wanted to get back into the space as applying machine learning techniques to the area of human health. At that point I didn't really know exactly what I wanted to work on. I looked around a little bit and decided to join Calico because first of all it's a life science company. It's a therapeutics company, in fact, within the Google umbrella, so I thought this would be a great place to integrate my interests of machine learning and human health. And also because I wanted to have the opportunity to work with amazing leaders like Art Levinson, who was the CEO who brought Genentech to greatness and [inaudible], who was the COO of Genentech and went with Art to start Calico, and I was really excited to have an opportunity to learn from such people.

DAPHNE: So I spent 18 months there and then came to the realization that the specific type of effort that I wanted to take on was much more of a foundational platform for drug discovery and not focused specifically on the problem of aging, which was Calico's charter. And so for me it made more sense to go and start that elsewhere where I would have more of the opportunity to tackle different therapeutic areas that might make more sense for such a platform without being restricted by the focus of what Calico specifically was interested in. And so I left Calico end of February of 2018 and started Insitro North much after.

CRAIG: I understand that drug discovery process using machine learning, it's an effort to narrow the field of possible molecules for testing. Is that right? You're not building molecules specifically for an outcome. You're looking at a range or family of molecules and narrowing the field to make testing less expensive. Do I have that right?

DAPHNE: Well, I think there are many ways in which machine learning could be applied and in fact is being applied in many different places in the drug discovery and development process. The area that you just talked about is certainly a focus for a number of companies where they're looking at the use of machine learning or more broadly computational models to identify which of a very large set of molecules might be potent binders to a particular target and then hopefully being able to screen a smaller subset than one would normally do. But there are many other places in which one could potentially apply machine learning in people who already are. So for instance, there's companies that are looking at identifying patients for clinical trial as ones that are most likely to be within the parameters of the trial.

DAPHNE: There are others that are looking at the actual management of the clinical trial and looking at biomarkers that are potentially more rich measurements than one typically looks at in most clinical trials where the patient comes in once every six months and a doctor or nurse evaluate their current state of health. But rather you can imagine interpreting data from wearable devices to try and see earlier which are the patients that are responding, which ones aren't. So there's even companies that are looking at machine learning in the manufacturing space to see how one can make manufacturing more efficient. We are actually currently at an earlier stage of this whole thing than even the molecules, which is identifying what are the right targets. Most drugs fail today because they're going after a target that doesn't meaningfully modulate the disease in a human.

DAPHNE: That's because it's very difficult to get data on what actually happens in a human being. And what we see today is that people often take a reductionist approach, which is they try and draw cartoon diagrams of pathways on the board and say, "Oh, this is a pathway that I believe is implicated in this and that disease. And look, this thing sits at the top of the pathway, so if I modulate that one, it's bound to have an effect." And it turns out that our simplistic understanding of biology, which is a really, really complicated space, is not usually enough to give us enough understanding of the disease that we can predict reliably what we can modulate that will have an effect, and furthermore, what we'll be able to modulate that doesn't have other deleterious effects on the human. So that's really where we're applying machine learning is on that side of it.

CRAIG: Can you give an example of a disease that you're focused on and how you narrowed down the targets so that you can rank them presumably?

DAPHNE: Yes. Let me start with the second half of your question, which is what approach are we taking to identifying targets and then we'll talk a little bit about therapeutic areas or indications. So the approach that we're taking is one that is driven by the growing realization that targets that have a basis in human genetics where nature has actually perturbed the target in different ways across different individuals, and we see an effect on clinical outcome that is commensurate without perturbation, targets that have that characteristic are much more likely to be successful in downstream clinical trials. We may not always understand why that particular target, when perturbed actually has the clinical effect that we see, but the fact is it's an empirical experiment that nature has performed for us that gives us empirical evidence that that target has meaningful effect in a human being.

DAPHNE: And furthermore, you can ask the question, does perturbing that target have deleterious side effects on people? Are there other maybe not so desired characteristics in humans that accompany that change in the target? That's a way in which a number of companies out there have now started to move towards the direction of looking at human genetics as a way of identifying targets that actually work in human beings. Not in cartoon diagrams, not in mice, but human beings. So that's the starting point. The next step though, which is a place where we're differentiated, is that we believe that while you get a lot of insight from human genetics, the studies that are currently available are often underpowered to really identify meaningful facts from the genetics alone and allow you to pinpoint which target is worthy of multi tens of millions or hundreds of millions of dollars of a clinical development program.

DAPHNE: So what we're doing is we're supplementing the human genetics data with a very large amount of in vitro data that we produce that perturbs different targets in systems that are derived from human biology, specifically from the STEM cells that you can derive from people with different genetics using the reprogramming work that was done by Yamanaka for which he received the Nobel prize. And you can now start perturbing targets and seeing what they do at the cellular level. And if you could use machine learning to appropriately interpret the cellular changes that you see when perturbing a target, you can now start to maybe narrow in on the targets that really do make a difference in human biology, and therefore narrow in on the targets that actually meaningfully make a difference in human clinical disease.

DAPHNE: So you combine the human genetics and clinical outcomes on the one side with a lot of in vitro data from the lab that is also human derived but at the cellular level and then you put those together to select the good targets. It's definitely not the case that all diseases result from a perturbation of a gene. In fact, very few diseases are solely the outcome of a perturbation and gene. Most diseases that we know of today, with the exception of fairly rare diseases with significant loss of function, oftentimes these are the novel variants that occurred during the creation of the first cell that is the fetus. Very few diseases that are just the result of the single gene. And the reason for that is that nature actually has a selection process against those. Most people with a really strong genetic disease don't procreate at the same level as people who are healthy.

DAPHNE: And so those variants get lost, and then that's why they're created de novo every time because there's this natural selection against passing those on to offspring. So most diseases are a combination of genetics, oftentimes in more than one gene and environmental factors. But that doesn't mean that one can't affect a meaningful change in the disease by perturbing a single gene and bringing it back to a healthy state. Maybe a poster child of that is the set of drugs that target the gene PCSK9. This is a set of drugs that are relatively recently improved in the last few years. They were discovered by human genetics. And what's interesting is that they were discovered by mutations in families where the gene was modified in one way and inherited. In this case, it gave rise to very high and relatively early onset in cholesterol levels, so it's called familial hypercholesterolemia.

DAPHNE: Interestingly, a different change in the same gene gave rise to the opposite effect of familial hypercholesterolemia where their cholesterol levels never went up no matter how much they ate burgers and fries and they were perfectly healthy and they very rarely got heart attacks. And so from these very strong mutations, there was realization that this gene could actually make a big difference to cholesterol levels and to cardiovascular disease. But what was also interesting is that even in the general population, when you looked again, you could see that people with small changes, much more subtle changes in the activity level of that gene also had a commensurate effect on cholesterol levels and cardiovascular disease. So even though the really, really strong, what's called highly penetrant mutations didn't occur often in the general population, much more subtle changes in the activity actually did occur.

DAPHNE: That's what gave people the hope that turned out to be true, that if you massively intervened at that gene and modified its activity to something that was more commensurate with what you saw in the highly penetrant version, specifically the hypercholesterolemia version, because of course we want to reduce cholesterol, not increase it, then that would actually be a therapeutic intervention and that turned out to be true. None of these are gene therapies right now for PCSK9, these are things that intervene by increasing or decreasing activities in traditional therapeutic modalities. So the first PCs canine drug was an antibody, actually two of them. The next one that's just been approved is an antisense oligo and that's an oligonucleotide based therapy that decreases the activity of the gene.

CRAIG: So beyond the genetic level, if you're looking at perturbations at the cellular level that's giving rise to disease, you were saying you were doing both or combining the genetic level and the cellular level can you give an example of how you combine them?

DAPHNE: Yeah. So imagine that we are able to take a set of cells, skin cells for that matter or white blood cells from patients who have a high risk for a disease and healthy controls who don't, and take those cells and using this technology that I mentioned earlier, what's called induced pluripotent STEM cells, which are these cells that you can generate from skin cells or white blood cells, but are STEM cells so that they can be then turned into any cell that exists in our body. And then now all of a sudden you can have standard cells, completely normal cells of whatever lineage is relevant for disease that come from healthy people and for sick people. And now you could look at them and say, is there anything that distinguishes them? Do they look different? Are the cells from a sick person different from the cells of a healthy person or are the people with a genetic variant that is disease causing cause they look different from the cells that ared derive from a healthy person.

DAPHNE: That gives you not only an understanding of what the disease looks like under a microscope, which is actually where the machine learning comes in, but actually a place where you could now start screening for interventions and ask yourself, is there a drug that if I put it into the cells, the sick person cells, they become more like the healthy person cells. So basically reverting the cellular phenotype is what we call it, that moves the cells from an unhealthy to a healthy state. That potentially isn't something that would be a clinical intervention that would have hopefully a meaningful clinical benefit to people with the disease. Because ultimately if we're able to make cells healthier, maybe we can also make the people healthier. That's sort of the basis for what we're doing and the human genetics comes in and how we take our understanding of the human genetic architecture of disease, and we create that in our in vitro system.

CRAIG: Have you gotten into clinical trials on any drugs or is this still in the research stage?

DAPHNE: I think it's important in general and not just for ourselves, for people to understand the timelines in drug development because I think there is a certain lack of understanding of that and that causes people to ask, well, no one has gotten an AI developed drug into clinical trials yet, so it must mean this not working. In the best of cases, getting a drug into clinical trials is a multi year effort. The time from ideation to approval on average is around 10 years. The time to get something into clinical trials, again on average, is about five to seven years just to getting them into clinical trials. So for most companies in this space, and I'm just talking for ourselves, this is just the beginning of creating a research engine that hopefully will allow us to get into clinical trials in some reasonable amount of time.

DAPHNE: But this is complicated because you're messing about with an actual live human and you're putting a drug in them that can potentially do harm, and so you want to be really, really careful before you go ahead and do that. So in that light we've been operating for a little less than 18 months and a lot of that was just even to get a lab up and running so that we can even start generating data far less interpreting the data and so no, we're not even close to clinical trial right now.

CRAIG: Yeah, and this process of identifying targets, is it also the case then that there are drugs in the market that are addressing disease inefficiently because they're focused on the wrong target and an even though focusing on that target has some downstream affects that are beneficial for addressing the disease, that there are other targets that are more directly linked to the disease? If no one's been looking at it using machine learning, could it be that there will be a wave of reassessment of drugs that are already in the market?

DAPHNE: I think there are absolutely drugs out there that are not targeting the optimal target. I think people have already found such examples even separately from the use of machine learning. I think equally so there is a very limited understanding today of which are the right patients to benefit from a particular drug. So I think that is perhaps an even more common problem where the focus, at least in most of the early years in drugs has been on let's find a drug and give it to everybody. And I think it's a natural inclination because you want to try and help as many people as you can. But the fact of the matter is for a lot of drugs there's a large subset of people that are not helped by that drug and there are even some for a number of drugs that are actively being hurt by the drug because maybe there's some aspect of that person genetic or otherwise that makes them particularly susceptible to adverse side effects.

DAPHNE: And there's really not been as much understanding of how do we appropriately target a drug to hit the exact right patient population, the ones that are going to be helped but not be hurt. So that's something that I think machine learning could potentially really help with.

CRAIG: And this process with the stem cells, how is machine learning applied in that case? When you're at the cellular level and you're looking for a target, what exactly are you doing? For example, in vitro, when you're working at the cellular level,

DAPHNE: Come back to the example that I mentioned before, which is you take these STEM cells, these IPS cells that are derived from people say with, let's take a really simple example maybe because I think that might help illustrate. Let's consider a scenario where you have a highly penetrant genetic mutation that has a very significant effect on people's clinical outcome. You could imagine taking a bunch of patient lines that have that mutation and a bunch of patients lines that don't have that mutation, those would often be healthy people. And let's imagine that you have let's say 50 lines or a hundred lines of each of them and now you convert them into the right type of cell because oftentimes the phenotype only shows in the right type of cell and then let's say it's a liver cell for instance, or a brain cell. So you take those STEM cells and differentiate them into liver cells or you differentiate them into neurons.

DAPHNE: And then you look at them using a bunch of the biological measurement approaches that have been developed over the last few years like for instance, RNA sequencing, which gives you a very accurate read out of the activities of different genes in the genome or you look at them under the microscopes and you measure cell morphology, like the shape of the cell and the position of the organelles and all sorts of other aspects of the cell. And with that set of tools you can use machine learning to say, okay, what is the thing, the patterns that distinguish between the sick cells in the healthy cells? Can I reliably reproducibly and in a way that extrapolates to new patients tell, the difference between sick and healthy. Now if I can do that, I can now ask myself, okay, I'm going to intervene in the sick cells and I'm going to put in something that increases the activity of this gene or decreases the activity to that gene.

DAPHNE: Or maybe even it's a natural product that I don't even know what it does, but I can put it into that cellular mixture and ask, does this drug meaningfully and reliably revert the phenotype from a sick cell to a healthy cell? And that is now something that if you see that happening repeatedly, you can ask, is this a starting point at a drug? And furthermore, you could also ask yourself for which subset of genetic backgrounds is this reversion happening? Maybe it's not everybody, maybe it's only people who have another gene in their genome that has a particular variant and it's only for those patients only on that background, if you will, that the disease phenotype is meaningfully reverted to healthy.

DAPHNE: I think it's that power to really look at the reversion of the phenotype across a diverse background of human genetics is first of all, what gives you the conviction that this drug is going to work and it's not just a fluke, but furthermore gives you the ability to judge which subset of patients it's going to work for so that you're not trying to treat multiple diseases with one drug. One of the biggest advancements that's happened in oncology in the last couple of decades is the realization that cancer is not one disease. So breast cancer for instance, is not one disease. There are patients that have had two positive breast cancer. There are patients that have Broca one defects. These are very different diseases and they manifest completely differently at the cellular level even though at the end of the day they all show up as a tumor growth in a breast.

DAPHNE: And the current drugs that are out there, the ones that are really effective are the ones that identify correctly what subtype of the disease does a patient have, and then you give that patient that drug and not a different drug. I think that is something we've not yet done for diseases other than cancer.

CRAIG: And going back to looking at individual differentiated cells from someone who has a genetic predisposition to a disease. You're building a data set then that a machine learning system can ingest and find features in, is that right?

DAPHNE: Yes.

CRAIG: And so on morphology for example, if you just take photographs of these cells and let the machine learning algorithms surface what features are distinctively [inaudible] predisposed, I mean you're letting the neural network presumably, I assume this is deep learning, surface those features that might otherwise be extremely difficult for a human to spot is. Is that one way that the machine learning plays a role?

DAPHNE: Yes, that is exactly right. What we take is oftentimes images from the cells and it's not just photographs because it's under magnification and oftentimes with particular stains that bring out certain aspects of cell morphology. Like for instance, you might stain the mitochondria or the nucleus or various proteins that are in the cell, which gives you an amazing richness of data in terms of the localization or even specific proteins within the cell. So there's diseases that might manifest in the translocation of a protein that's supposed to be one place and all of a sudden it's somewhere else, or two proteins that are supposed to be next to each other cause they're supposed to be working in concert and now they're no longer as next to each other. So there's things you can see under the microscope that are often very significant to human disease and are quite subtle.

DAPHNE: So by giving the machine all of this high content data and asking it, what do you see that distinguishes between the healthy and sick? It oftentimes find some pretty subtle patterns that in our experience, even a trained cell biologist does not necessarily perceive.

CRAIG: So that's the target identification aspect, simplified understandably.

DAPHNE: Well, it's a way of identifying a phenotype and now you're looking for target as something that reverts that phenotype from the unhealthy to the healthy state.

CRAIG: I see. But for example, you were talking about placement of particular proteins. If you identify the phenotype and in that phenotype there is an anomaly in the way that the proteins are placed and that's consistent across cells of people who are predisposed to the illness, how then do you identify a molecule that could change that?

DAPHNE: Well that is the next stage of the process and there's a number of different things that one can do. One can go and do what's called phenotypic screening, which is something that a good fraction of the drugs that are out there were discovered by phenotypic screening, which is you take a collection of drugs and you just try and see which are the ones that seem to be going in the right direction. When they succeed you say, okay, that's the starting point for us to do what's called lead optimization. So the first drug that you find is called the hit and then you do what's called hit the lead where you try and improve the properties of the molecule to make it more like a highly potent drug. That's one approach and it starts out just purely with a phenotype. A different approach, which is something that is much more accessible to the world today is, well, once you have a phenotype, you can look for a genetic intervention using CRISPR that modifies the phenotype and reverts it.

DAPHNE: Now that may be something that you would then target using genetic therapies, but mostly you wouldn't. You would just say, oh, if I perturb this gene, it reverts the phenotype, now let me look for a drug that perturbs that gene. And now that gives you the opportunity to identify drugs often in a more scalable way by looking for molecules that bind to that particular target. Now by doing so, they may or may not actually move the phenotype in the right direction. Maybe they bind to the target, but they mined in the wrong place. Or maybe they even do the wrong thing, but it gives you a smaller subset of molecules that at least in our case you can output back into that same phenotypic assay as we call it and ask whether those drugs that we discovered as being binders to a particular target actually move the phenotype in the right direction. So it's a way of doing a much faster winnowing down, if you will, of the space of molecules to something that is much more likely to be relevant to the target. So they find out the [inaudble] and then you test those.

CRAIG: Presumably there are many different strategies for identifying likely molecules. Is one-- If you're talking about binding is one the shape of the molecule. You've got neural net and you want it to identify molecules that are likely to have an impact, one of the things that you're looking for, is it the shape of the molecule, because binding is the lock and key metaphor.

DAPHNE: Yeah. Once you have a target in place, there are a number of different approaches that one can use to identify potential binders. There are methods that you might think of as just group four screening. You have a library, most pharma companies for instance have libraries of two, three, four million compounds that they basically just have them sitting in little vials and you have the protein and once you have the protein, if you're able to purify it, you bind it to a hard object and then you flow molecules over it, and then at the end you see which was actually bound. So that's one approach. A second approach, which is what's called DNA encoded small molecules, they have libraries now of a billion molecules that each has a DNA tag on it and then again you flow those molecules over the target and at the end of the day you look to see which are the ones that bound by sequencing the DNA tags that are stuck on the protein targets.

DAPHNE: Those are some experimental approaches. There are also computational approaches that do this by effectively looking at, as you said, the small molecule structure. You have some notion of the protein shape that you might get from extra crystallography or something and then you basically computationally dock the molecule into the protein to see, as you said, the lock and key structure. That is something that requires a much deeper understanding if you will, of the shape of the protein as opposed to the more experimental methods where you could potentially employ them even without having a full understanding of the structure of the protein. Here you need to have some understanding of the structure and again, you now look at potentially billions of molecules in this case and see which of them might fit the shape of the pocket as it's called in the protein that's called virtual screening.

DAPHNE: It's something that has been very much used in the drug discovery space by pharma companies and others for decades now, a couple decades. I would say that this has been utilized quite effectively. So you have physical screening technologies and you have virtual screening technologies. The application of machine learning to accelerate that process is something that I think is now starting to emerge. I think it's early days in the sense that there's not been as of yet sort of a real breakthrough result in which a machine learning model has found a molecule that has been elusive to a person. I think there's progress towards that goal and I'm hopeful that in the next years we will find-- a computer will be able to identify molecules faster and in some cases even in situations where a person might struggle, but I think that remains to be proven.

CRAIG: You were talking about looking at existing databases of molecules and using machine learning to identify those that are likely binders if you're looking at that aspect. Is it also possible to use machine learning to create molecules?

DAPHNE: It is in fact possible. There's been a number of publications in the last I would say two years, that have used ideas like what's called GANs- Generative Adversarial Networks to actually generate molecules directly. As I said, I think at this point today and it can change, it remains to be proven that the process by which the computer does that actually is more effective than that of what a medicinal chemist does. That a computer does it faster or is able to do it in cases where a person would struggle. I'm not saying it can't be done, I'm just saying it has not yet been done.

CRAIG: There are several other companies that are pursuing similar machine learning strategies. Do you feel like this is the future of drug discovery or has it yet to be proven?

DAPHNE: Well, so first of all I want to just respond to the first part of your statement, and again just to clarify, there is a broad space here and machine learning is an incredibly powerful tool. Drug discovery is a very complex process. There are many ways in which machine learning across its multiple variants can be applied in very different ways to different stages of the drug discovery process that are in no way remotely similar to each other. Sometimes people tell me, "You know, there's this and that company and it's applying the machine learning to drug discovery. Are they competitors?" It's like, no, they're applying natural language processing to understanding text abstracts and good for them, but it's completely unrelated to what we're doing, other than it uses the words 'Machine learning' and "Drug discovery' in the same sentence.

DAPHNE: I mean this is in no way saying that what they're doing is not useful, but it's just completely different to what we're doing. So with that, I think important thing to keep in mind, I think as I said a minute ago, machine learning is an incredibly powerful tool. Some people will say, "Well, it's going to be just like x-ray crystallography and drug discovery. It's going to be a small little tool that helps us in one aspect of drug discovery, but it's not going to transform the space." I think that's not perceiving the full impact that this technology could have because it could be deployed so broadly across so many different parts of the drug discovery and development pipeline. I also don't think it's a silver bullet in the sense that it's not going to be that we just wave our magic wand and machine learning to the rescue is going to solve all of the things that make drug discovery really hard.

DAPHNE: Drug discovery is genuinely really hard. As I said before, you're intervening in human life, in human health, you have to be really, really careful. Understanding of the biology is very partial and we keep discovering new, amazing things that we didn't know about that can sometimes completely shift the picture. So it's not something that one undertakes lightly and it's not going to be, as I said, this fairy dust that magically transforms what is a very hard problem to a very easy problem. So I try and take this nuanced view that says, yes, it will be increasingly important in this space. It will make a difference, but it's not going to be something that happens instantaneously.

DAPHNE: And really important is the fact that in order to apply machine learning well in this space, you really need to have the people who understand biology and drug discovery on the one hand and the people who understand machine learning on the other work effectively as equal partners on a team, rather than having little silos of people doing machine learning who think they understand biology or little silos of people who understand biology who think that if they just throw something into a convolutional neural network, magically it's going to work. You really need to have people who think about the problems together and that's what we're trying to build in Insitro.

CRAIG: That's it for this week's podcast. I want to thank Daphne for her time. If any of you want to go into greater depth about drug discovery, you can download a transcript of this episode from our website Eye-On A.I, that's www.eye-on.ai. I encourage you to take a look at the Infinite Red's mini-course on machine learning at www.learn.infinite.red.

The singularity may not be near, but AI is about to change your world, so pay attention.