EPISODE 1112

[INTRODUCTION]

[00:00:00] JM: Drug trials can lead to new therapeutics and preventative medications being discovered and placed on the market. Unfortunately, these drug trials typically require animal testing. This means that animal are killed or harmed as a result of needing to verify that a drug will not kill humans. Animal testing is unavoidable, but the extent to which testing needs to occur can be reduced by inserting machine learning models, which simulate the effects of a drug on the human body. If the simulated effect is negative enough, animal testing does not need to be run. Thus, no animals will need to be harmed.

Bryan Vicknair and Jason Walsh work at VeriSIM Life, a company which makes software simulations of animals. These simulations could be used to model drug testing and change the workflow for drug trials. They join the show to talk through the mechanics of drug testing and how VeriSIM fits into that workflow by simulating a human and being able to run simulated drugs through that human.

[SPONSOR MESSAGE]

[00:01:05] JM: Operations teams can find themselves choosing between two options, either take full control over the infrastructure yourself, or give all developers permission to access production. The first approach increases the operations team's workload, which often results in overwhelming situations and ops becoming a bottleneck. And the second approach allows for rapid deployment of changes to production, but it causes a serious risk to infrastructure uptime.

With Octopus Deploy, operations teams have a third option. The Octopus platform can be authorized to run approved steps and play an intermediary role. Octopus delivers self-service without sacrificing control over production, and it also provides a comprehensive audit log of the changes that are being made. Developers can enable self-service with automation. By automating the processes that are forming a bottleneck, developers can free themselves from the waiting game. You can learn more about Runbook Automation at octopus.com/runbooks.

Octopus can help you with your Runbook Automation and just go to octopus.com/runbooks to learn how.

[INTERVIEW]

[00:02:17] JM: Bryan and Jason, welcome to the show.

[00:02:18] BV: Thank you, Jeff.

[00:02:19] JW: Yeah, thanks for having us.

[00:02:21] JM: Today we're talking about drug development. Describe the process of developing a drug for humans.

[00:02:27] JW: So, developing a drug has couple major stages, but it generally goes through drug discovery where they find a whole bunch of compounds. And then drug development, where we generally deal with where you take those compounds and whittle them down to the compounds that will go through clinical trials, etc. So there are a lot of in vitro testing, animal testing, etc., along the way, but those go along in defined stages. And ultimately you end up with a drug in clinical trials that may or may not be approved for the market.

[00:03:03] JM: And there is this process by which drugs can get filtered out of the candidate testing process. So you can think of this as a funnel. If I'm trying to cure COVID-19, at the top of the funnel, I have every single drug imaginable. Maybe I have worm root extract. I have Advil. I have all kinds of things. And as we try different things, things are getting filtered out. How do drugs get filtered out through the candidate process of being a drug that can cure something?

[00:03:36] BV: There are a couple ways in which that happens. Certainly, early on in the process, you do a lot of chemical studies to see whether something is going to hit a certain target. And then you move into the testing phases in drug development where you're trying to find out is this drug going to be toxic? Is this drug going to sit at the concentration level that it's going to be effective? Is it actually affecting the body in the way that we expect it to? That's what happens in those testing levels. It's more and more safety testing. More and more efficacy

testing as you go along to bigger and bigger animals. Once you get into humans as well. It's all different levels of that.

[00:04:18] JM: How well do animals serve as model organisms for how a drug will impact a human?

[00:04:24] JW: I mean, I think it's safe to say that it's pretty poor model actually. In general, 92% of drugs that fail on humans – I mean, 92% of drugs that show success in animals end up not being effective in humans. That's pretty bad odds, right? That's 8 out of 100. We can do better than that, and that's one other thing that via cell, our technology at VeriSIM Life is trying to improve, is that preclinical phase, right? Not only do we want to stop the animal testing for ethical reasons, right? But it's also very expensive and a very long process to do this animal testing. And the results just don't scale well to humans, right?

One of our major contributions to this problem is that our technology replaces the animal testing with in silico methods. We can use computation to predict these outcomes and we can be a lot more effective than this animal testing.

[00:05:20] BV: Absolutely. And to give a bit more context into just how sometimes they scale those, if we're comparing our complex computational methods versus just taking a rat model and scaling it up by bodyweight to a human, which sometimes happens. It's the amount of complexity you're able to factor in just goes off the charts once you can actually get these kind of simulations into play as supposed to saying, "Ah. It should be okay in terms of weight."

[00:05:50] JM: So there's a diagram on your website that illustrates what you're trying to do with VeriSIM Life comprehensively, and it illustrates this funnel that basically there is no getting around the fact that taking a drug through the testing process is really time consuming. It's really resource-intensive and it has the unfortunate side effect of probably killing a lot of animals or torturing a lot of animals. And that's just the way things are. We've decided that the human life is incrementally more relevant than an animal life or perhaps that animal testing is just necessary to save some greater number of lives, some greater good.

But you have this visualization where you basically show that with VeriSIM, you can reduce the number of animals that are getting testing on. And we will eventually get into software-related things. We just really have to tee up why this technology is useful. Can you explain the bottlenecks in the testing process that you're focused on with VeriSIM Life?

[00:06:52] JW: For sure. We very much focus on that segment that is lab testing, animal testing, etc., because as Bryan said, you have that not only five-ish percent of drugs that pass animal testing are actually making it through the clinical trials. Our focus, and one of the things that got me very excited when I joined up about this being something that we can really drive forward in the industry is that taking that 5% up to, say, 25%, 30% of drugs passing animal testing going to human – Passing in humans, is the difference of millions of dollars in savings and the difference of far less animal testing needing to be done. And it can be a lot more targeted and you can say, "Okay. Test these drugs this way on these animals and have it be a lot more focused. Save a lot of those sort of scattershot animal testings that happen and save a bunch of time that is time to market that sick people really need.

[00:07:53] BV: And to follow up on that, to give your listeners some idea what this is solving, right? To get a drug to market, it could take 10 to 12 years and over a billion dollars. So there's a lot of room for improvement there, right? If you improve both of those numbers by 30%, it's huge savings just for the human race. That's where it's super exciting to work for those companies, because there's an ability to have a very real impact on human health. It's one of the things we love about working here.

[00:08:25] JM: There are two terms that I'd like to define for the sake of this conversation, pharmacokinetics and pharmacodynamics. Could you explain what those two terms are?

[00:08:38] JW: Yes. Pharmacokinetics is – And we can break this down very simply. Pharmacokinetics is very much what the body does to the drug and to see how it behaves through the bloodstream and different organs, etc. And then pharmacodynamics is what the drug does to the body. What effects you're seeing in certain places as it goes through. There's an interplay between the two when you're studying things. We focus a lot on pharmacokinetics in the past. But there's a lot of room in both to be done in computational spaces.

[00:09:11] JM: Okay. As we get into the software side of things, there are two types of models that we are going to be discussing in this conversation. There're models of organism and then there's machine learning models. And I can explain how a machine learning model works. I can explain how a model organism works, but I'm sure you guys could explain those in more detail. Could you break down the models? What we're talking about with models as we are exploring VeriSIM Life?

[00:09:41] BV: Yeah. Like I said, there are two major approaches, right? There's the mechanistic approach and the machine learning approach. The mechanistic approach, we're going to use that when we understand the relationships of inputs to outputs in some system. If we understand the cause and the effect to some degree, we can model that with mathematics and we're going to create this mechanistic model.

For example, we can replicate an organ if we understand the relationship between the initial dose of a substance. How it flows to organ through the blood, the volume of the organ, absorption rates, etc. Right? Humans understand the system, and we attempt to create some simplified model of it. And that will give us the ability to predict results.

The machine learning approach is slightly different. In the machine learning approach, we can predict outcomes of biological systems using statistical learning approaches. But we never really attempt to understand the mechanisms of the system. It's good enough for us in ML to just make predictions about how that system is likely to behave. That's the major difference between those two approaches.

[00:10:50] JM: Okay. So let's drive this home. If we are trying to replicate the effects of a drug in silico, as it were. The computer simulation of how a human response to a drug, how do we do that? How can we replicate what a human would do or how a human body would react to a drug?

[00:11:16] JW: At VeriSIM Life, we come at it from both sides. We have mechanistic models and we have ML models and they both inform each other. But in general, if you're going to simulate what a drug does to the body, a key part of that is modeling how that compound is going to flow through the body. How it's going to be absorbed. How it's metabolized and how it

eventually leaves the body, right? So we need to – Just like in software engineering, we have a big problem. It's hard to solve a big problem, so you break it into smaller problems. We do the same thing when we're modeling the human body. We don't model the whole body in one go. We're going to create separate compartments. In each compartment might be an organ or a set of tissues, right? And then we can focus just on that little model, and it's really a function. It has inputs and it creates an output. It's really like a state machine.

So you get several of these compartments together. Maybe one is the heart. Maybe one is the kidney. One is the liver. And then you compose those together, just like we do in software, right? You compose those together in a larger model that sort of dictates where the inputs and the outputs are flowing to between the compartments. And that's how – That's a high-level view of how you can model something as complex as the human body just by breaking it down into smaller problems. And that's the mechanistic approach. The ML approach looks similar, but you're not modeling why things are happening in ML approach. In the ML approach, you're getting predictions out of it.

[00:12:51] JM: Right. I mean, the thing that deep learning is good at is taking a lot of unstructured data and reasoning about how similar unstructured datasets will trigger responses in the future, or how they will be categorized or whatnot. But I want to take the hypercritical journalistic approach here and just kind of say like this seems too unfathomably complex to model the human body. Like, "Come on. Seriously? You can't do that." How would you respond to that kind of naïve, like I just don't believe it type criticism?

[00:13:31] BV: That, from a complexity perspective, we can see exactly what you're talking about. And really, we model to the point that it will be useful for the outputs that we're looking for. And what that means is we take the pieces of the human body, equations based on established science, put those together and say, "Okay. We understand how the general, say, blood flow in these situations is going to be." And we understand that this this will cover us down to a certain precision. And we need this precision for what our customers are looking for. And we need to be able to say this is what the drug is doing.

Yes, we're not accounting for what every single cell in the body is doing. We abstract out to a certain level as all models do, but we take it down to a level enough that we know what the

broad strokes of the drug in the body are and take that to a very useful outcome that we're looking for. Maybe that is concentration of the drug in the body. Maybe we need to predict what certain parameters are going be in the drug. But the goal, we need to focus on the endpoint, and that will tell us the precision that we need in the model.

[00:14:47] JM: And I guess it's worth noting that it's not like you throw these experimental model drugs into a computer model and then say, "Okay. Pass us the computer model. Let's start a phase of testing in humans." It goes from in silico to animal testing, right?

[00:15:09] JW: Yes. And one of the things there is, one, we validate every simulation that we perform. Whenever we do that, we're often validating our results against established data in terms of – Again, I mentioned blood concentrations, if we've gathered items there. We are predicting those blood concentrations at any given time. We predict with accuracy metrics exactly where that lies and what percentage accuracy that we're getting.

We'll be able to understand roughly where we're at accuracy wise. And yes, then that goes into the more targeted animal testing that we were discussing saying, "Okay. We've whittled down many more drugs than we would have using scattershot animal testing, and now you can verify that these much more focus group into that stage of animal testing.

[SPONSOR MESSAGE]

[00:16:08] JM: This episode of Software Engineering Daily is brought to you by Datadog, a full stack monitoring platform that integrates with over 350 technologies like Gremlin, PagerDuty, AWS Lambda, Spinnaker and more. With the rich visualizations and algorithmic alerts, Datadog can help you monitor the effects of chaos experiments. It can also identify weaknesses and improve the reliability of your systems. Visit softwareengineeringdaily.com/datadog to start a free 14-day trial and receive one of Datadog's famously cozy T-shirts. That softwareengineeringdaily.com/datadog. Thank you to Datadog for being a long-running sponsor of Software Engineering Daily.

[INTERVIEW CONTINUED]

[00:17:01] JM: All right. Let me just ask another kind of naïve question. You've got these in silico models and you've got a drug. Is there like a method call you have that's like eat the drug now? Like model.ingestdrug? Ingest some number of milligrams of this drug?

[00:17:22] BV: Yeah. There is. I mean, there's almost a one-to-one correspondence there. Let's look at that at a high-level, right? Say we're going like simulate the liver. If we think of the liver, we can think of it – If we're going to model it, right? Let's treat it just as a finite state machine. Okay? And we just need to understand how that liver model, how that function, how that state that's in RAM somewhere goes from T0 to T1. We just need that step function. So we can model this sliver with a function called step liver, let's say. And what does that function do? It takes the current state of the liver and then it gives you the next state of the liver in reality, in time. That's a high-level view of how you can model that.

Now, what goes on in that function? That's the hard part, right? That's our team of modelers, and computational chemists, and scientists. They are the ones really informing the software engineers how this should happen. So it's not on the software engineer to figure out how to model that. We work with experts in disparate fields that help us out with that part. Our job is to implement it, right? At a software level, yes, you can think of these compartments really as pure functions. That's how they can be implemented. Of course, there're all sorts of logging and stuff like that. But at a high-level, yes, that's how it is done. It can be done.

[00:18:38] JM: Can you share an example? Do you have a drug that has gone through VeriSIM life that was proven to be nondestructive or was proven to be nondestructive enough to move on to an animal testing or is this all confidential? Do you have anything you can share?

[00:18:55] JW: Ultimately, the specific customer projects that we have in place are confidential. So we can talk about them in broad strokes. But when it comes to specific outcomes, that's mostly confidential at this point.

[00:19:07] JM: Okay. Can you give maybe the category at least? Like a cancer drug or some like that?

[00:19:12] JW: Our model takes all sorts of drugs. We try to make this as drug agnostic as possible. So when it comes to the types of drugs that we've done, we've done studies focusing specifically on transdermal drugs, meaning being absorbed through the skin. We're trying to do all our part to make sure that we're contributing to solutions to the current COVID-19 problem. And we've put out notices about that.

I mean, it spreads to basically wherever our customers want to go. We can model that and tailor our approach to the problem that they're trying to solve. Really, what we do is we start with the problem and we build everything from there in terms of our modelers, in terms of our scientists, and then product that we as software engineers put together.

[00:20:02] JM: Can you tell me more about how the software model is architected? Where do you even begin?

[00:20:13] JW: So we begin really with our core simulator. As Bryan mentioned, this is a step function that just goes through maybe millisecond after millisecond, or whatever time step that we set that to. And within that, we have the equations that our modelers put together, and those drive the behavior of the drug in different parts of the body. And all of that comes together to reporting on – Often, we're reporting on what's in the bloodstream. It all comes together in one cohesive simulation that then you're able to graph and see, "Okay. That's what the drug was doing through the body."

If you want to zoom out from there, we do a number of other things in terms of breaking out the parameters that go into these simulations. Every drug has particular parameters, and sometimes we have all of them. Sometimes we don't. And that's where a lot of our machine learning approaches come in to predict the rest of these parameters that go in based on the data that we've seen in the past.

There are things like that. And then there are more specialized modules that we build from the modeling space where they say, "Okay. We need to insert more complexity into how a drug is ingested. We need to insert more complexity into how something clears out of the body." So it starts with that core simulator. We have things that fill in missing data. And then, really, it's just building complexity on top of that.

[00:21:45] BV: Let me zoom out a bit just for your listeners' benefit, right? We're all computer scientists. We think in terms of inputs and outputs. At the high-level, what does all this mean? We really want – The inputs are the drug and maybe some properties of the body. Are the disease? Are they not? And what's output going to be? In a simple simulation in a lot of cases, it's a time concentration graph. We really want to see at each step in time how much of the drug is in the body. And in particular, how much of that drug is in each compartment in the body? Those are the inputs and the outputs for this problem, right? It's a simply stated problem. It's just massively complex to actually implement it.

[00:22:24] JM: And I still don't understand how you gauge accuracy and like what the feedback loop is. Because I just imagine, you put this thing and it's like, "Okay, we think we have ruled out drug. We think we have ruled out drug B. We think drug C works. We think it doesn't seem to cause like liver problems, or intestinal tract problems. Move it on." How do you validate that?

[00:22:53] JW: So the validation comes in, in what Bryan was talking about with the time concentration curves. We have empirical data that we've gathered from public sources that are given to us in partnerships, etc., and we make sure to validate our models, our entire system against what we already know to be true. So we'll run those drugs with exactly the same – With the same parameters, with the same set ups against what happens in real-life in terms of either the results out of a human, or our model can extend animals as well. So maybe we've measure that against an animal too.

But once we get those curves next to each other, we have a whole slew of accuracy metrics that we get to say like, "Okay. Here is where we're potentially going wrong here, here," and that just allows us to tighten the model further and further such that we have more and more confidence in its output. And we've already taken that to a significant degree.

[00:23:54] BV: And so this this validation, process that Jason just described. Through this process, we've identified a bunch of – Like holes and inconsistencies in this scientific data, right? Because when it comes down to computer, it's got to be right or wrong. There is no gray area. So when you're putting this data into a database and through an ML pipeline and through these simulations, it becomes very evident, like where there're mistakes in the originally

published paper. And we'll actually feed that information back to the people, the groups that publish it so that they can make those corrections.

But we have just microscope upon microscopes on this validation data. Because like you said, Jeff. You have to have confidence in these results. So we have an entire team of people basically dedicated to this. Looking at this data, right? Getting more validation data. Getting better validation data. So we do that.

Another thing we found is that holes in scientific empirical data that's available, we can use machine learning methods to fill those holes, right? Of course, that whole process is validated. Validation results in the data that lies inside of our database is a huge effort that we're constantly working on. Remind me again, you lean on the structural model or you lean on machine learning models? Or, sorry, it's some combination of those two?

[00:25:16] JW: It's absolutely a combination. We essentially start this, we'll say the skeleton of it is our mechanistic model that is informed by the biology. We have the modelers that put together the equation side of it. We have scientists who verify the veracity of what's getting put together. So it's a very interdisciplinary effort just to get that into place. And then the idea of integrating the machine learning models in there are saying, "Where are the holes in the process? Where are the black boxes that we have to understand?" Sometimes that is the missing parameters that we've discussed. Sometimes it's just another part of the process that we just don't have enough scientific understanding and we need to make sure that we're measuring that purely on outcomes, as Bryan said. It very much is a mix of the two, and that takes a lot of interdisciplinary communication. And that's part of the reason why we are the first people to really be tackling it in this manner, because in a lot of cases, you have these large disparate teams with each specialty in each team and you don't have the focused agility of bringing everyone together and say, "Let's do a single effort and try to bring all of our understanding into this combined model approach."

[00:26:33] JM: What parts of the drug testing process can and cannot be simulated?

[00:26:41] BV: Well, that – I mean, we could talk about theoretically what could be simulated. Let's go there, right? What can we simulate mechanistically? What we mostly tackle, the system

has to be deterministic. If it's indeterministic, then I don't know. Roll a die or something, right? We mainly focus on systems like that, systems where we understand the physiological processes. We're able to model anything like that, finite systems, right? We use finite computational resources. So we need to focus on finite problems. But practically, there are limitations also, right? So, computational resources being one of them. We run everything in the cloud and we rent a bunch of space. But you only have so much money, so much time, so much space. Just like anybody else in this field, right? We run up against that. These computations take a long time. We have to run just hundreds of thousands simulations to get results. We run up against that problem.

From the machine learning side, what can we simulate? What's reproducible? Really, anything that we have enough data about, right? We need a lot of data and we need good data. And that's really the bottleneck on the ML side of what can be simulated. If you give us enough data and it's good data, then we're going to be able to give you good predictions. And the synthesis of these two things, the mechanistic model and the ML approach, that's the key. That gets us further. And then we can simulate more systems than either system can simulate just by themselves.

[00:28:13] JM: And does this whole thing sit in-memory or is this a distributed system? How do you manage that?

[00:28:20] BV: It's distributed, and it's mostly has to be in-memory just for performance reasons. But, I mean, we have a lot of data too. So that's going to be living on not spinning disks anymore, but it's living on disks. But we try to keep everything, yeah, in-memory as much as possible when you're doing these simulations.

[00:28:35] JW: And one thing to talk about the distributed system, one thing that we really put a big effort on last year is to make sure we could scale up these amount of simulations as much as possible. So we're not trying to run these hundreds of thousands of simulations at the same time on a single box. We made sure to utilize clustering mass parallelization and just a number of other techniques to scale ourselves up to running hundreds of simultaneous simulations at a time. Each of which don't necessarily take an incredibly long time, but if you're scaling up to the

total being, a million for a single large report that you're trying to compile. It became necessary. So we had a large effort toward that last year.

[00:29:19] JM: And what programming language do you use to build a model of a human?

[00:29:25] JW: Generally, our models are built in Python. We have, for performance reasons, we've made sure to do the core part of our simulator. They're doing most of the floating-point operations in Cython to give us significant performance boost. But Python is the core language of most of our simulations, internal tools, etc. We're mainly a Python job.

[00:29:49] BV: And so something about that too is that the modelers that we work with, the computational chemists that we work with, they're not programmers. But a lot of them have exposure to programming via Python. It's kind of like the lingua franca. So that's a key reason we use Python. It's something that modelers can write in. They can read and write it. So they can express an algorithmic idea in Python and hand it off to us. And that's the way we communicate. And then we do the software engineer thing and make it robust and production-ready and do our thing on it. But a lot of the people in this company can read and write Python.

[00:30:25] JM: The Cython case is interesting. You're saying you only use Cython for some parts of the model. Cython being that trance piled version of Python that is in C, or C++, I think. And so why not just put the whole model into Cython if it runs faster?

[00:30:48] JW: Because only some parts of the model are incredibly computationally-heavy, where we're trying to make sure that we're putting the most effort into what will give us the best return. So when it came to looking at our codebase and seeing what was worth, making sure it could run well on Cython and tying to that sort of compile cycle, because that is something that you introduce when you're trying to run things on Cython. We found that this one core area was where we would get the best benefit out of it and we really did see that just getting things on to Cyphon there in the right way gave us a three times boost in throughput. And so focusing on that area really gave us some good dividends.

[00:31:36] JM: Okay. Can we revisit the end-to-end workflow of running a drug test? Tell me again, what happens when the model consumes a drug.

[00:31:48] JW: When the model consumes a drug, and you have to understand that consuming can come in different ways. We're talking about you could injected a drug. You could eat a drug. You could do a number of things in terms of how you introduce it to the body. There's that piece of making sure that the model starts taking in the right amount of drug over time so that it either takes it in all at once. Or gets it spaced out. And some of it is potentially lost. Say, you take it orally in the digestive system.

It goes through that process. You end up with what starts circulating around the bloodstream, and then takes through those time steps based on that quantity in the bloodstream at any given time. Simultaneously running with that, you might be adding more through that ingestion process or whatever else. But then as soon as we start the simulation, it's just going to the next time step with whatever's available, and we'll make more available over time based on what we expect to happen by that absorption. And then we just go through and understand like where, in what sections of the body things are going to be cleared out and in what way the drug concentration is going to be affected. So it's really just understanding out what points and in what way that drug concentration will change.

[00:33:09] JM: And can you tell me a little bit more about the software stack? What requirements do you have? Does it run in Amazon Web Services? Do you use any other random platform tools, or SaaS tools, or distributed queuing systems? Give me a little bit more of what's in the software stack.

[00:33:31] JW: Yeah. As we said, Python is the core to this. And on the machine learning side, we use a lot of specific machine learning Python libraries based a lot around Pandas and using TensorFlow, etc. But then getting out of the Python space, we do base a lot of our cloud computing in AWS. We use the basic tools there, but some of the more specific tools we use for massive scaling type work. We dig deep into ECS and SQS, the queue system. And I would say, otherwise, tools on the frontend, we use a lot of React, a lot of JavaScript in terms of making sure that our entire customer presentation is very smooth. Bryan, what do you have on a daily basis?

[00:34:22] BV: Yeah, mostly that. I'm either writing Python. I'm messing around with React and Redus. Working in RDS instances, right? AWS's implementation of MySQL. Those are the main tools that we're hitting every day though.

[00:34:37] JM: Zooming out a little bit, can you tell me about the go-to-market strategy? How do you find, I guess, drug companies? That's the customer, right? A drug company?

[00:34:49] BV: Yes, that's a drug company. That is our customer.

[00:34:51] JW: And it's actually one of the potential customers. So we'd spread who we're looking to with, who we're looking to have as a customer throughout a number of different places in the sort of biotech and pharmaceutical space where basically anyone who is trying to understand more about a drug, anyone who is trying to develop a drug, we are open to. So we deal with academic institutions, with larger drug companies, maybe with biotech startups.

This scope with who we look for from a business sense is pretty wide. And we really enjoy having that sort of validation for the model in all these different places, because we've had a lot of success in branching out like that.

[00:35:34] JM: Are there any other – What are the other kinds of customers for VeriSIM other than other than drug companies?

[00:35:40] JW: Mainly talking about the like academic institutions and such who are doing research and maybe partnering with drug companies. But it's more than just working with the large company drug companies.

[00:35:53] BV: For example, like COVID, right? There could be academic institutions that we're helping with our platform to do research on COVID. And the funding might come actually from the government on that one, right? Not from a large pharma company. It might come through a grant.

[00:36:09] JW: And there are a number of other situations where we work through grants as well as supposed to – And sometimes that's in partnership with an entity, sometimes that's ourselves. But that's part of what we do as well.

[00:36:19] JM: Are there a lot of competitors? What is in silico drug testing market look like?

[00:36:27] JW: There's some level of competition and that there are computational services that allow you to like work through and build up a single model through doing a lot of work on your end and making sure that you are doing all this testing. There is some that they use in the computational sciences space to understand what is going on with a drug. No one is doing what we are doing now. We are taking anything like that and moving it to a much broader space. No one has generalized this system like us and are able to say, “Look. You give us your inputs. We have this model set up and maybe we’ll be able to customize to some degree for you to account for some different things.” But we’re taking a lot of the work out of this and say, “Hey, we have a validated model that is able to give you some insights into what you're looking for. Let's make this process a lot simpler for you.” There’s no one else to our understanding that’s doing that.

[00:37:34] BV: And a lot of the software that is available, it was written in the 80s. It’s like mainframe stuff. Or it was written in the 90s and it's like just – It’s classic 90s UI. Nothing web-based. No ML pipeline informing anything like that. We see the pharma industry in general, we see it as being underserved by modern software technologies and delivery systems, and that's one thing we’re excited to do, is kind of bring it up-to-date.

[00:38:02] JW: And one of the challenges that we have is just making sure that we’re things the right way and just saying that, “Look, we have this.” And being able to differentiate in communication between us and anything else that they've used in the past, because we’re such a step change that it requires a shift in understanding to be able to say, “Okay. This is how this is going to help us and how much of a golf this is going to narrow in order of our abilities to develop drugs.”

[SPONSOR MESSAGE]

[00:38:40] JM: If you listen to this show, you are probably a software engineer or a data scientist. If you want to develop skills to build machine learning models, check out Springboard. Springboard is an online education program that gives you hands-on experience with creating and deploying machine learning models into production, and every student who goes through Springboard is paired with a mentor, a machine learning expert who gives that student one-on-one mentorship support over video.

The Springboard program offers a job guarantee in its career tracks, meaning that you do not have to pay until you secure a job in machine learning. If you're curious about transitioning into machine learning, go to softwareengineeringdaily.com/springboard. Listeners can get $500 in scholarship if they use the code AI Springboard. This scholarship is for 20 students who enroll by going to softwareengineeringdaily.com/springboard and enter the code AI springboard. It takes about 10 minutes to apply. It's free and it's awarded on a first-come first-served basis. If you're interested in transitioning into machine learning, go to softwareengineeringdaily.com/springboard.

Anyone who is interested and likes the idea of building and deploying machine learning models, deep learning models, you might like Springboard. Go to softwareengineeringdaily.com/springboard, and thank you to Springboard for being a sponsor.

[INTERVIEW CONTINUED]

[00:40:17] JM: Very interesting. Can we take a step back? Tell me how you see the healthcare landscape evolving over the next 3 to 5 years, or maybe it’s a 10-year time horizon. How do you see drug development changing?

[00:40:34] JW: I see drug development going a lot more toward this sort of in silico validated space, because we've already seen the FDA starts to warm up to these kind of approaches. And that was something that was very exciting for us. It’s pretty obvious why. But we see this time-to-market shortening. And exactly how long that takes is very much dependent on how quickly the market is able to adapt to these source of technologies and weave that into the pipeline. But if we’re able to – And people seem to be very receptive toward what we’re trying to do. If people are willing to bring these computational tools in line with their process, then we’re talking about

knocking down that 10 to 12-year drug pipeline down by maybe only order of years and maybe just over the next 5, 10 years.

Again, that depends on rates of adaption. But we're already seeing due to COVID, due to major world events, that it is a lot less viable to be doing some of these things in-person. To be doing the massive amount of in-person testing that has to happen for a drug to get into play. If anything, world events have tended toward pressure on the industry getting toward these in silico methods.

[00:42:01] JM: And how does that change the overall market for pharmaceuticals? Does it mean that we get the same types of pharmaceuticals on the market? We just get them faster? Does it mean that we can develop entirely new drugs that we wouldn't have otherwise? What is the actual ramification?

[00:42:21] BV: We're going to get drugs faster and we're going to get drugs cheaper and we're going to get better drugs. While we're all sleeping, the models are going to be doing the number crunching. This time span, this 10 to 12 years to bring a drug to market is going to shrink. We're going to get drugs faster. And COVID is just a great example of everybody wants that faster, right? We don't want to have to wait through all the preclinical trials and the clinical trials. Yeah, were going to get them faster. All of this is much cheaper. Animal testing, huge, large clinical trials, these are all massively expensive things causing over billion dollars to bring a drug to market. So that cost is going to shrink, right? We're going to get them cheaper. Yeah, we're going to get better drugs too.

Because just like ML does in other areas, the computer is going to start pointing out things that we weren't even looking at, and it already does that on our platform, and it starts to narrow. It starts saying to us, "Hey. Stop looking over here. You're never going to find something over here. You start looking over here." And then you can take the domain experts and send them that way. And so, yeah, we're going to discover drugs that we wouldn't have discovered otherwise.

[00:43:30] JM: Do you have any other predictions for how machine learning will improve healthcare?

[00:43:36] JW: I think there are a lot of opportunities for machine learning in terms if we're staying in sort of the drug space, personalize dosing is something that is very popular in health space. And that is it's something that's there's plenty of scientific literature behind that. But it has to adjust very quickly to adjust to variables in a care setting. So that's something that I've sort of been tracking and it's something that obviously being adjacent to this space we pay some attention to.

It's something that if solution comes on the market to be able to dose drugs for a patient based on all the conditions of their care, that's going to drive some great outcomes. And I think there are some people already studying those approaches.

[00:44:25] BV: I think there's something else. I hope that it's going to happen, I think well, is that, in general, there's going to be a basic machine learning literacy that's going to permeate society as it just continues to be a thing on our lives. And I think that's going to inform some of the choices that the scientific groups who are doing experiments make. I think we're going to see a shift in how they present their data, how it can be reproduced. I'm hoping that data becomes more parsable, right? Because we spent a lot of effort right now going through just old papers and grabbing data.

But as soon as general society start realizing the benefits of ML, which I think that they do, even though they don't know what machine learning might be. They're all benefiting from it. As soon as they know that, "Hey, this ML thing can get us some good stuff." And if they know just a cursory glance of what it need to be successful, that will inform their decisions in what they're doing. And so I'm hoping that the data that flows out of places becomes more parsable basically to machine learning.

[00:45:30] JW: And one of the things driving that very much is the interoperability of data that people like subscribe to just common forms for their data to take as it's being transitioned from company to company or whoever is holding it. I actually came up in the healthcare tech space in Epic systems during the period of time that they were very much focusing on interoperability. And that is a Herculean effort. So one of the things that it's definitely worth paying attention to is seeing the efforts that happened, the coalitions that form around making sure that healthcare

data, research data is able to be ported from place to place. Because the more that we're able to do that, the more ML solutions are going pop up and really be able to make a difference in the healthcare space.

[00:46:23] JM: When you are at Epic, Epic gets this really bad rap. Epic is a medical or EMR company, right? It's like this super integrated EMR system. There actually has a really interesting story around it. I remember reading some really long article in like New York Times or Atlantic or something about the idiosyncratic founder of Epic. So that brings to mind that like the question of why is EMR the way it is. And is it ever going to change?

[00:46:48] JW: I would say the change is incremental as it is in many cases in the healthcare industry in general. As software engineers, we are so used to things moving quickly. And in the healthcare space, in the biotech space, things do not move that quickly. Change happens in a much more conservative manner. And that's why like we're trying so hard to make the biotech space catch up to the understanding that we have right now to the capability that we have in our technology. Because it's not just a matter of creating a good product. It's creating a good product that your users are going to want to use.

In terms of like here at VeriSIM, when we're presenting something, we need to make sure that the scientists on the other end are going to want to use it. Does that mean we have to hold back on maybe some of the more interesting technologies in flash of things that we could put into place on the frontend that may be would make life easier for power users? Sure. But we also have to think about the core users that you're talking about in the space.

There is a bit of a lag time there. And I think the industry, the healthcare industry, the biotech industry at large is trying to minimize that gap and trying to bring things up to date. But there is always that push and pull in that space.

[00:48:21] JM: There is this speculative hypothetical technology of this this toilet that we will all have some day, where it's a toilet where every day you use the toilet and it tells you what is going on inside of your body. Are we ever going to get that toilet?

[00:48:38] BV: I hope so. That sounds great.

[00:48:42] JM: How far are we from that technology?

[00:48:43] JW: There are so many different things that go into that where – And some of the things that obviously fall into our wheelhouse in terms of simulating exactly what's going on in there with the inputs and outputs that you're given. So when it comes to like providing useful simulations in those sort of context, we're already a lot of the way there. But then one of the big issues there as we've discussed for VeriSIM's product already is integration.

One of the things for our company is we have the technology and we need to make sure that it integrates correctly with the science. that all the insights are there together. Same thing happens with that toilet. If you're talking about whatever inputs that toilet is giving. Well, you need the people who understand those to come together with the people who can simulate what's going on with the body, people who can put together the correct aesthetics. You're now talking about an appliance. Really, one of the limiting factors there is getting all of those different people to come together to make that product.

[00:49:49] BV: I think, also, we're all software guys, right? My opinion, like that's a hardware problem. Good luck hardware guys, basically. Get the nano part, get the nano machines going, right? Let's inject it in us, and then just give us the data. And then the software guys will take over, basically.

I've been curious about how a society going to react to, let's say, you can have implantable chips. Let's say you can have injectable nano machines that could report all these things. What's that gap between the Fitbit and the injected nano machines? There is a psychological gap there. That's going to be huge, I think, for some people. I am not exactly sure where I fit on that right now, but we'll see.

[00:50:29] JM: I guess one other question. We've done a few shows about these cloud labs. There's like Emerald Cloud Lab and Transcriptic. Do you guys have any take on the cloud laboratories and their viability in improving experimentation and drug development?

[00:50:48] JW: When you talk about Cloud Laboratories, what specific approach are you talking about?

[00:50:53] JM: basically the ability to call out to a web service and have it run some kind of system that needs a PCR machine or some other set of machines. Basically, the democratization of the hardware for laboratories.

[00:51:10] JW: So we're talking about these sort of in-person tests, but we have central labs that are executing these tests for maybe like a smaller drug company or something like that.

[00:51:20] JM: Yeah. Yeah, or an individual.

[00:51:22] JW: I mean, it's sort of another – In some ways, another form of what we're trying to do in terms of making testing more accessible. I think that it's going to be necessary. And if they are able to standardize things to a point where they are the gold standard, then maybe that is the core validation testing in animals that couples with the in silico testing that we provide in the first place. So maybe in that sort of situation, you would say, "Okay. We'll help you narrow it down way better than you can do in the industry right now." And then these people have been proven to have the best lab set up for what you're trying to do. They have their – Well-rated by all the different companies that have worked with them. And then they can bring that then in that part of the pipeline. So it's slightly different parts of the pipeline depending on what they want to use in terms of lab testing and animal testing. But it's definitely an interesting way for the industry to go.

[00:52:32] BV: Yeah. That's fascinating. I mean, I see that as just like instead of mining Bitcoin, just all night long, just be running – Just be messing with petri dishes. And that would be great. If it's all automated and if it had an API and I can just tell it, "Yeah, I want a thousand of these little tests run overnight," and then it just generates the data. I mean, to us, that's what we want. Give us that experimental data, because our software just feeds off of that.

So if we had a pipeline, an automated pipeline that is getting real empirical results, that sounds awesome. And the fact that if had an API, we just love that stuff. So yeah, I think that would be great.

[00:53:11] JM: Last question. Do you have any predictions for how VeriSIM is going to look in 3 to 5 years, your company?

[00:53:20] JW: Oh man! We are scaling so fast and changing it in many, many different ways. It's a very interesting question and it's one that we're very, very excited about. I mean, in terms of actual employee growth, we see that expanding significantly. But we see ourselves becoming – I was talking about the gold standard in terms that in terms of physical labs, but we want to become the gold standard when it comes to this sort of in silico testing and be a go-to to say, "Hey, take all this work off of our hands." So that means building out a number of different products. And we're working on making our product as modular as possible to be able to handle as many different problems as customers can give us.

Really, what we see is something that we're able to say to a customer, "You have these particular goals when it comes to what you're trying to find out in your drug pipeline. Great. We have this and this and this part of our process. Maybe we give you some custom development as well in there." But the goal is that we just have our model validated to the point that we're able to proclaim complete and utter confidence in as many realms as you can think of and expand out to all the different parts of drug development that you can imagine and just really become the go-to-space for drug companies to try to speed up their process.

[00:54:55] BV: Yeah. I think that what we're going to see over the next 3 to 5 years is the industry is going to become comfortable with these in silico methods more and more. And our core team right now, we're a super close knit team, super cross-functional. We know it works. We touch it every day. It's incredible the technology that we have right now. It's really just getting the word out there and making people comfortable with the methods that we're using. That's one of the big challenges, right?

So a lot along the technical problems are solved. We're doing it today. It's happening. After this podcast, I'm going to do it the rest of the day, right? So we just need to get the industry comfortable with these technologies, right? And we have a team to do that. We have leadership to scale this company. I mean, five years is even kind of crazy from anything. In three years, this company is going to be completely different. It's just going to blow up.

[00:55:53] JW: And that really just relies on momentum and really just getting us – As we were saying earlier, that sort of translation of expectation that we bring to the customers and say, "Look. This is what we can do for you. This is how your process can completely change for the better. And the more companies, the more organizations that are able to recognize that, the more of that momentum that we have. Really, over the next three years, being able to keep up that momentum will greatly transform both this company and the drug space in general.

[00:56:25] JM: Guys, thank you so much for coming on the show. It's been a real pleasure.

[00:56:28] JW: Absolutely. Thank you.

[00:56:27] BV: Thanks, Jeff. Appreciate it.

[END OF INTERVIEW]

[00:56:38] JM: When the New Yorker magazine asked Mark Zuckerberg how he gets his news. He said the one news source he definitively follows is Techmeme. For more than two years and nearly 700 episodes, the Techmeme Ride Home Podcast has been one of Silicon Valley's favorites. The Techmeme Ride Home Podcast is daily. It's only 15 to 20 minutes long and it's every day. By 5 PM Eastern, it has all the latest tech news, but it's more than just headlines. You could get a robot to read you the headlines.

The Techmeme Ride Home Podcast is all about the context around the latest news of the day. It's top stories, the top posts and tweets and conversations about those stories as well as behind-the-scenes analysis. Techmeme Ride Home is like TL DR as a service. The folks at Techmeme are online all day reading everything so they can catch you up. Search your podcast player today for Ride Home and subscribe to the Techmeme Ride Home Podcast.

[END]