Rockin' HIT Sales

Episode Transcript

Standards That Scale: Building Healthcare AI That's Testable, Reliable, and Safe

Guest: Ram D. Sriram, Chief of the Systems and Software Division, National Institute of Standards and Technology's Information Technology Laboratory (ITL)

Release Date: April 22, 2026

Transcript edited lightly for clarity and readability. Intro and outro omitted.

Ram D. Sriram ***(00:01)***

Thank you David for having me here and I'm ready right now to answer your questions.

David Hacker ***(00:07)***

Okay, well, Ram, thank you very much for joining me on Rockin' HIT Sales. You're at the National Institute of Standards and Technology. You work at the intersection of trustworthy AI and standards. So today our conversation is gonna help our listeners translate what that means in practical guidance for digital health teams building and deploying AI. in those real world clinical settings. So you've been working in AI since the early 1980s, correct? So you lived through the knowledge-based systems, the neural networks, and now the large language models. When you look at today's health care AI tidal wave, is there anything that jumps out that feels genuinely new?

Ram D. Sriram ***(00:38)***

Mm-hmm, you're right.

David Hacker *(00:55)*

And conversely, is there anything that just feels like an old story that's repackaged or rebranded?

Ram D. Sriram *(01:02)*

Yeah, I mean, as you know that there are three waves, they call three revolutions of AI. We went through the knowledge-based revolution, then we went to the neural network, a deep learning kind of thing, and now we are seeing the neuro-symbolic aspects of it. Now, if you look at it in a 1997 book that I wrote, I talked about all three revolutions, where I talked about knowledge-based systems, neural networks. And now neural networks can be combined with knowledge-based systems. And the title of the book was Intelligent Systems for Engineering, a Knowledge-Based Approach, because I was working in the engineering field at that time, before I came to NIST. This is when I was a professor at MIT. OK? So and if you go back and see and look in the current environment of large language models and so on, really all we are seeing right now is this realization of this predictions we made about 30 years ago. So large language models are really, in fact, in a 1986 video, which is on the web, I talked about knowledge in your pocket. Okay, so this was that time, that time there were no smartphones, internet was sparse, they used to use it for email and so on. But again, I envisioned about 20 years from then, there will be this knowledge in your pocket. And we're kind of seeing that in a way with these LLMs, where I can take my cell phone around here, get on to either Gemini, chat or Claude, whatever your favorite system is, type out something about my symptoms and I can get some diagnosis. So we are seeing a realization of these things around here, but in different format. What I envisioned as a knowledge-based expert systems, now we are seeing that in a different format, like the LLMs or the neural networks. But these things right now are going to hit a wall very soon, so we are going to go into the neuro-symbolic kind of reasoning. Does it answer your question?

David Hacker *(03:00)*

It does, thank you and and yes going back to 86 you had quite the correct prognostication as to where AI was going to go. So when you talk a lot with NIST, you you state that first off that at its core that NIST is a measurement in standards lab. And so for our listeners who may not understand this definition, what does metrology for AI mean and why does it matter so much to them?

Ram D. Sriram *(03:31)*

Yeah, actually you have to go back to the notion of metrology as such, that's all measurements. So if you look at National Institute of Standards and Technology, it's both measurements, standards, and we also have a technological focus so that we want to have a major impact on the industry to kind of research that we do. So essentially our goal is to improve the US industry in the global markets and to help the US industry compete in the global markets. That's the kind of mission we have been given. And we are also part of the Constitution Article 1, Section 8, where they talk about weights and measures and so on. Now, in the past, we have been mostly in terms of the physical standards, like when you go to a gas station, you you put gas, how do you know that's one liter or one gallon, it's traceable to NIST. So these are all physical standards. But now we are in the era of the information age. In the information age, we've got to measure information, so that an AI is a kind of information.

So how do you measure the AI systems? And that's part of our goal around here. And also, the AI systems are becoming very popular, as you know. Like, now if you talk to your mother, grandmother, they're all using ChatGPT, or one of those chatbots they're using. So you can see that it's pervasive right now. Everyone is almost everyone knows about AI at this stage. So when we do that, it's becoming part of our lives. So we need to have a mechanism to measure these things too.

David Hacker ***(04:35)***

. Okay.

Ram D. Sriram ***(04:58)***

And that's what I meant by metrology for AI. But at least we also use AI for metrology because we are a measurement lab and we want to know how accurate those measurements are. We use artificial intelligence for doing that. So it's metrology for AI and AI for metrology.

David Hacker ***(05:15)***

Interesting, interesting. And kind of sticking with that measurement, if we think about AI as an instrument, if you will, like a lab device or a scale, what are the basic questions that should be asked about how it's being measured and being calibrated before that AI can be trusted, especially in the clinical environment.

Ram D. Sriram ***(05:38)***

Now, it's kind of an interesting question because when you talk about devices, you actually have the software devices and the hardware devices in the medical field. So as you know, the AI systems are really software devices in a way, but they're also embedded in the hardware devices. Okay, like when you have your pacemaker, that's actually going to be a very intelligent system in the future. Similarly, your CGM, the continuous glucose monitors, even your Apple watch, all these things are really smart systems with software. So that's a very important part to understand the context around here and also what's the purpose of this instrument. What are you actually using this instrument for? It's an interesting part of it. And also, and then the measurements depending on what instrument you are using. So if you're using an electronic health records, your set of metrics are different than you are using a medical device that measures something in your body, a body parameters as such. So. And variable devices are another important part of it, as you know that they're both hardware and software embedded in here. And so there's a lot of things going on and we need to investigate metrology for AI in all these things. So, but then in terms of calibration, when people talk about calibration, we also have to talk about ground truths. Okay, what is the ground truth here, especially when I talk about a user calibrating? Is that clear?

David Hacker ***(07:07)***

Yeah, very, very. Thank you. Thank you. So I'm going shift gears for just for this next question. I think it goes back about six years ago. You proposed the idea of the P9 medicine. And that means that care that is personalized, predictive, preventative, participatory, pervasive, precise, privacy preserving, protective, and priced efficiently. Now, for those listeners that are building out a digital tool or AI product, how do you feel that those nine P's should influence their product roadmap? And again, going back to measurement, what should they measure for those nine P's?

Ram D. Sriram *(07:51)*

Yeah, that's an interesting question because if you look at the nine piece part of it, that kind of evolved through some talks with my colleagues, Sab and Jane over the few years, it was based on Leroy Hood's P4 medicine. In the P9 medicine, let's take for example one, it's kind of hard to talk about. Everything and trustworthiness as we are going to talk about in the future will play a major role in all these things. But if you look at trustworthiness as such, one of the things in the P9 part of it is the protective. So how are you going to protect and privacy? So you got to have measurements of privacy and protectiveness. And then the other part of it is, for example, predictive. How are you going to predict that you're going to have a particular disease or not? And all these things, the user has to have trust in that. Trustworthiness becomes extremely important in all these things. So if you look at the metrics, trustworthiness is one of the metrics. Similarly, accuracy is another metrics. So if you are talking about a system which I'm giving you diagnostic output, I want to know... How accurate is this term? How precise the system is? So in terms of precision, you want to know how accuracy becomes very important in this precision. So for each of this P9 medicine, I can map into a metrological aspects and interoperability. And I talk about pervasiveness in this thing, that is the medicine should be given anywhere, anytime, we're wearing this in variables, which are talking to one another and interoperability becomes a big issue around here. So then you get into semantics of this interoperability and you got to measure. You got to make sure that information that is sent from one device is accurately replicated in another device.

David Hacker *(09:59)*

That leads me, that's a great lead in for my next question because in the past you've talked about how the smart network systems and societies were sensors, devices. EHRs and even social networks all contribute data. Now, from a developer standpoint, what does that mean to design for that rather than just using it using just a single hospital apartment or a single data source?

Ram D. Sriram *(10:27)*

yeah, so you are getting data from different sources as you can see. Like for example, my Apple watch is constantly monitoring my heart rate. So that heart rate and in fact sometimes ECGs and all those things are given out by this Apple watch. And similarly, I might have another device which is monitoring my sugar levels, like the CGMs, the continuous glucose. glucose monitors is measuring that. So each of these devices are monitoring various things. In fact, it's interesting in the future, the food you eat also is going to be monitored. Okay, to find out whether it's good for you or not so good for you. All those things are going into something called the personal health record as we talked about in my talk previously on the P9 medicine. So the issue then here is that you have different data sources. So one has to have to integrate all of them in a proper manner. So we need appropriate standards for integration of these things. Okay, standards become a very important part of this whole process. So if you are a young startup company or you want to get into this thing, my first advice is to go for the low-hanging fruit, okay? And see that if you can, some or other, see what are the... Already available standards for these things and how we can use them to effectively integrate. Don't try and invent new things. Try and see what is out there and try to put them within your framework. And one of the things that you need to think about carefully are use cases.

David Hacker *(12:06)*

Okay. Now, NIST has published an AI risk management framework with ideas like govern, map, measure, manage. At a very practical level, what is measuring AI? I'm going keep going back to measuring AI, but what is measuring AI look like for a healthcare use case?

Ram D. Sriram *(12:25)*

OK, so again, a little bit background for the readers about the NIST AI framework, the AI RMF has got four major modules, the governing part of it, the mapping part of it, the measuring part of it, and the managing part of it. So these are the four elements of the NIST AI RMF. And any of these organizations, whether it's health care organization or someone doing financial transactions, the homeland security, and those kinds of things, will have to take this particular if they are using AI tools. In fact, any software tools is applicable too, because as you know, AI is a kind of a software with certain nuances in there, particularities for AI, depending on what kind of AI that you use. So all these things are very important when you are doing a kind of application and we're using AI. Now, one of the things that could potentially be is a scenario where an hospital is using an AI tool. to detect some disease like pneumonia or something like that from chest x-rays. So once this particular thing is that, and if you're going through all the four stages like the governing part of it, where the hospital establishes a multidisciplinary deem and put all the rules in place, and then mapping is all to do with context, and now we are coming to measurements. That's a question that you asked. And in terms of measurements, one of the things that the hospital should do is to develop a set of metrics to monitor whether there are false positive rates or whether false negative rates and those kinds of things have to be measured. So this is just one example of measurements. you're getting, so you're using X-rays and the system kind of diagnosis saying the person has got some lung cancer or something like that, okay? So once it does that, then you want to know whether then you have to have some kind of a matrix that this particular system that they're using in the previous past. what is, how many false positives it had, how many false negatives it had, what are the type one error, type two error, those kinds of things will have to be put in place. So the matrix is trying to figure out how accurate the system is essentially. So it's one of the metrics. There are all kinds of other metrics in there.

David Hacker *(14:37)*

Okay. Now, if I'm getting ready or I have built or in building the roadmap for a new clinical AI, and you've seen a lot of different products, is there anything that you or your teams see where the companies are under investing in today? Things like is it data quality, semantics, standards, security, privacy, or maybe something else?

Ram D. Sriram *(15:08)*

Yeah. So as I mentioned in my recent blog, which is available on the web for the NIST, if you Google my name and NIST and AI, standards are often overlooked in the development of these AI frameworks. And also, the other thing that one needs to ask is, what is the key purpose on what exactly we are planning to achieve? For that, I would recommend Simon Sinek's video if you have not watched it on purpose. And most of the people kind of miss out the purpose. What's the purpose? Why am I doing this in the first place? Why? What? How? We need to answer these questions.

David Hacker *(15:51)*

OK. Now, repeatedly in your talks, you use a phrase, uncertainty quantification. Can you help explain that to us non-technical people, why knowing when a model doesn't know is just as important as its overall accuracy?

Ram D. Sriram *(16:10)*

Yeah, so on this thing, let's take an example of medical devices. Yeah, good example. So let an AI enabled medical devices. So what happens is that if like, let's take go to an hospital. In the hospital, a patient is sent in as an oncologist if his biopsy is diagnosed as malignant, okay? So that means the patient comes in. you take in the device, whatever device that you're doing like MRI or X-ray or something like that, tells you that this particular patient has that particular tumor in there is malignant. Now, it can also say that that particular thing is not malignant. It's benign. is that what happens in all these cases is that if the AI system is, let us say that then that particular AI system has been trained with a particular set of data. So the AI system now will have to diagnose properly. How do you trust this in the first place? There's uncertainty involved in this diagnosis because your particular X-ray may or may not be in the X-rays that it was trained for. So it's out of distribution. So then you have to make a decision. Should I send it this biopsy to an oncologist? Sorry, biopsy to a pathologist so that the pathologist can tell me if it's benign or if it's cancerous and then goes to the oncologist or through the system can it directly go to the oncologist, not to the pathologist? If I'm uncertain, if my uncertainty is, if I'm a little bit uncertain here, if I'm not really confident in the system, I first send it to the pathologist to make sure that this particular thing is 100% accurate so that the oncologist can come over. On the other hand, if I'm 100% confident in this particular device is giving the right things within my distribution, then I can send it directly to the oncologist. I don't need to go through this pathologist path. So do you see in this case, either I go directly to the oncologist, and if I'm uncertain about this, I go through the pathologist. and from the pathologist to the oncologist. So, uncertainty comes. So, now I got to evaluate the uncertainty of my devices.

David Hacker *(18:41)*

Thank

Ram D. Sriram *(18:47)*

And that's a whole different ballgame of how to do this uncertainty. And at least we are doing some work on that area. How do you characterize this uncertainty?

David Hacker *(18:56)*

Now in the past, you've identified that setting up standards, if it's done properly, that that can actually accelerate innovation, not slow it down because sometimes there's friction between standards and regulations and innovation itself. Can you give me an example, anytime in your career where a good standard or benchmark actually did accelerate innovation?

Ram D. Sriram *(19:18)*

Yeah, there are several examples that I can give within my career at NIST. The cloud computing is a good example where because of the standards that we have developed, can actually, previously there was no common understanding what's a cloud, what the cloud did and the stakeholders were

all confused. And once we develop the standards, everybody can follow the standards. so innovation in cloud computing actually improved drastically. and the people came up with all kinds of new things. And the other thing is in terms of the product modeling, we developed a standard called CPM and OAM. And from that, we developed something called OntoStep, which is we don't need to know the details. But then it so happened that these particular standards were used by the industry to improve actually their product development process. So they ended up in better designs and better design manufacturing capability. because we developed this interoperability standard. So that's a second example I can give you. And the third one is in the healthcare field, which is that we do actually testing of the interface between the EHRs and the immunization state registries. And because of that, of our testing, the quality of the data actually improved. because the data which was of poor quality, for example, was thrown away and data of the good quality was sent across. And by establishing this common language and high data integrity, we have actually sparked a new era of innovation in this area. And in fact, actually during the COVID-19 time period, our systems actually helped a lot in making sure that the right data was sent to the right place. without any problems.

David Hacker ***(21:15)***

Yeah, and I'm sure everyone is thankful for that. Because that was a difficult time, that's for sure. And thank you for those examples. Yeah. Now, a lot of our listeners work for, you know, startups and growth stage companies. And

Ram D. Sriram ***(21:20)***

And this played big role.

David Hacker ***(21:31)***

As a smaller organization, feel they really can't influence the work that's being done on standards. Is there a realistic way for these smaller companies to engage with the standards and measurement efforts that you guys are doing?

Ram D. Sriram ***(21:44)***

Yeah, actually, apparently there's, in fact, it's a good question you asked, because I've been thinking about that too. There's something called the NIST public working groups, because these NIST public working groups, they allow the smaller teams to contribute without needing to have a membership in a big standards organization. So one needs to look more into the NIST public working groups.

David Hacker ***(22:09)***

Okay, great, and certainly put that out there. Now, we're about ready to wrap things up. I have two final questions. I refer to them as my lightning wrap, two quick, hard-hitting questions, very simple, hopefully. The first one is, what's one question every company should ask themselves before they put an AI-enabled feature in front of clinicians or patients?

Ram D. Sriram ***(22:33)***

I think one of the most important questions that this company should ask is that, how can I convince my doctor or the patient that my system is trustworthy? And give an example that like so many times when you go through these LLMs and you put your symptoms and it kind of gives you

some diagnosis either from a patient perspective or a doctor's perspective, which may or may not be right. That's a fundamental problem. The patient is always worried about, I trust this in the first place? And you, for people who are using chatbots, they'll see that many times it gives answers they just cannot believe.

David Hacker ***(23:18)***

Okay. And my final question, is there one red flag behavior or with AI or data that makes you a bit nervous when you see it out in the field?

Ram D. Sriram ***(23:31)***

Well, we talked about these things called trojans. And sometimes it happens where people manipulate these AI systems. And so as a result of that, you could potentially come up with a diagnosis which was not meant to be there. And so you have to be very careful and you have to look at it. And that's the reason why I talk about neuro-symbolic computing. So these systems should have guardrails. That's the most important thing. Having the guardrails, is very, important in these systems.

David Hacker ***(24:04)***

OK, Ram, thank you so much for joining me today and for making the standards and that trustworthy AI conversation practical for our listeners to understand and comprehend. So again, thank you very much for joining Rockin' HIT Sales.

Ram D. Sriram ***(24:20)***

Again, David, thank you very much and it's been my pleasure. And you also have given me some food for thought.