Better understanding through sound statistics

Transcript

Dr. Peter Pronovost

The output of every AI model or LLM model is some form of a probability statement. It could vary, but it's essentially a probability statement. And there's compelling and overwhelming data that clinicians are horrible at interpreting probabilities. I mean, absolutely horrible. They … we make bad decisions all the time because we don't understand it.

And there's been some research, not in AI, but just in general. Much of it comes … a great book called *Risk Savvy*, that comes out of the Max Planck Institute, showing that if you show physicians graphics like of 100 little stick figures and you shade in what would be the equivalent of the probability, they make much better reasoning, than if you say, "OK, this has a 50% probability, this has a 30(%), how do I make a decision?"

And I think there's been way too little research, or even just evaluation, of, "How do I present the output and how do I build trust so clinicians will use it?" Because I was part of the development of the TREWScore. That was probably one of the most, you know, impactful sepsis scores. And it, it has like, AUC (area under the curve) of, like, 0.94. I mean, it performs really well. But none of the clinicians would use it because it was like this black box. And they … they're making a clinical decision that's life or death. And they were … there was no work to trust it.

And I think it's never going to be acceptable to say, "Trust the model because it's my license down the line," or, "It's the patient's life on the line," that we're going to need a set of rules about how these models perform that say, you know, "What did I test it? How did I perform, who did I do it on?"