Ruth Adewuya, MD (host):

Hello, you're listening to Stanford Medcast, Stanford CME's podcast where we bring you insights from the world's leading petitions and scientists. This podcast is available on Apple Podcast, Amazon Music, Spotify, Google Podcast, and Stitcher. I am your host, Dr. Ruth Adewuya. This episode is part of our Hot Topics miniseries. In this episode, I'll be talking to two experts about the impacts of ChatGPT both on learning and in clinical medicine. I started my conversation with Dr. Victor Lee.

Dr. Victor Lee is an associate professor in the graduate school of Education at Stanford University and leads the data interactions and STEM teaching and learning lab. He earned his doctorate in learning sciences at Northwestern University. Through his research, he asks what future-facing STEM knowledge, tools, and practices are important to know to enable active participation and critical engagement with our increasingly digitally infused lives. Dr. Lee is also the faculty lead for the accelerator initiative on generative AI in education. Dr. Lee sits on the editorial board of several leading journals, including the Journal of the Learning Sciences and Cognition and Instruction. He's a past president and elected fellow of the International Society of the Learning Sciences. Victor, thank you so much for chatting with me today.

Dr. Victor R Lee (guest speaker):

Thanks for having me.

Ruth Adewuya, MD (host):

I want to start with basic definitions, we've heard the terms ChatGPT, AI, AI language model. Can you start with defining what is ChatGPT?

Dr. Victor R Lee (guest speaker):

ChatGPT would technically be considered a chatbot, and other versions or other types of chatbots that you may have heard about are Google's Bard or also Microsoft Bing, which uses similar sorts of underlying technology. But the whole premise is it's technology that you can talk to by typing and it gives very relevant seeming responses.

Ruth Adewuya, MD (host):

Our conversation today is really to focus on AI technologies like ChatGPT and its implications within education. How can AI technologies like ChatGPT be utilized in education?

Dr. Victor R Lee (guest speaker):

I would liken it to the way in which the internet and search engines have been it could be utilized. We're looking at a new generation of technology, but it basically provides access to large amounts of information and in a form that comes off as conversational. So that's a big difference with respect to how things are with something like Google and just running a regular search on it. So because of that, you have all this information available, but just like with Google or any other internet search, you're going to have to approach it with some skepticism and know that there is some bad content that is out there, and that even though it may sound very conversational and present itself as seeming knowledgeable, it doesn't really know what it's saying. It just knows words that go well together and it comes across sounding as very expert like.

So if you wanted to get new information, if you wanted to check some information and have it told back to you in a more accessible or comprehensible form, if you wanted it to help you generate things that

you had to spend a lot of time writing such as example cases or documents or letters, it provides a very reasonable starting version of that upon request.

Ruth Adewuya, MD (host):

And I have had the opportunity to play around with that idea, if you input some information for an email and it can generate some things for you. I think one of my own personal experiences was I'm currently in a master's program and I remember getting an announcement from them about how Turnitin can now detect ChatGPT as part of its plagiarism detection services. I'm curious to get your thoughts on that given how you've defined how AI can be useful in education.

Dr. Victor R Lee (guest speaker):

We're beginning to see more instances of incorrectly identifying cheating, so saying somebody used AI generated text when they actually did not, and that is one of the very serious concerns that should be considered in the use of it. But also with technologies like Turnitin or other things that attempt to be AI detectors, AI text detectors, it can be outsmarted. They are trained in a similar way to other sorts of large language models, and so they really rely on what information it's been given. So if you give it something a bit different, it can be thrown off. So what it means then is we need to think a bit more about why are we concerned about these notions of cheating and what would it mean if we restructured the educational experience so that way that's not the main focus, that we don't make our determinations on how thoughtful people are based on how well they can write a particular form of testing.

Ruth Adewuya, MD (host):

I really like looking at this as an opportunity for educators to take a step back and really think about what success is for the student and what is the grading process and what is cheating. So I really appreciate that context. Another terminology that I've heard has to do with artificial hallucinations, which obviously in the clinical world, hallucinations means something completely different. First, could you define what that is? And then the follow-up question is how can this phenomenon impact the integration of ChatGPT into medical curricula or usage for medical students?

Dr. Victor R Lee (guest speaker):

Hallucinations, similarly for when humans have hallucinations, we consider it as describing or presenting something that by our metrics of reality was not really there. So an AI hallucination would be generating incorrect information. In the academic realm, if you ask for references for certain research studies or papers, it does make some up, because the model is very good at structuring things in the way that a reference would look like and putting in the appropriate names that would be associated with that type of clinical work or some of that research, but it will completely create papers that have never existed. And so hallucinations are a name for that phenomenon and this persistent risk of misinformation or inaccurate information.

I do have concerns, as do others who are a bit critical of AI, about this notion of hallucinations in that, again, this is a very fancy pattern matching machine, it's not truly thinking, and so if we expect it to judge correctness or factuality, that's really not what it is built for. With respect to medical education, what we need to really think about is how widely known are these hallucination phenomena and these misinformation and accuracy concerns amongst all the people who could be using it, and are people being vigilant with respect to checking and monitoring that and keeping everyone accountable for going

to really trusted expert sources who have human insight and knowledge and experience that an AI technology just simply can't.

Ruth Adewuya, MD (host):

That's helpful to know that it's not an accuracy machine, and I think that whenever we have conversations around AI and ChatGPT, there's always this concern about it taking over the role that humans play. And what I heard you say is that there will always be a role for human moderation in this process because it is a pattern recognition system.

Dr. Victor R Lee (guest speaker):

Absolutely, and thinking about medicine as a context, it's a very human practice. It's humans trying to help other humans using the best knowledge available and the best recommendations possible. That's fundamentally human work that we do. AI is a nice enhancement, but it's not a replacement per se. It'll change the things that we need to do with respect to the amounts of it or how much we need to be concerned about it. So the analogy I would give is we now have auto-correct and spell checking technology on our mobile phones in our Word processors that make it so that many of our typos and errors can get corrected.

Of course, there are some very funny mess-ups that come from that same technology, and so when we realized that the text messaging system changed what it is that you wanted to say because it did not recognize the word, we need to have that oversight, but it means that also we don't have to spend extra time and energy in large amounts making sure that we've spelled everything perfectly while we're composing it. So that's a change that we're looking at, but you still need to have the human oversight because it's something that helps humans to do what humans do, but it's fundamentally the humans who are doing that core work, which is expressing self, communicating, forming relationships, et cetera,

Ruth Adewuya, MD (host):

That's a really helpful distinction, and I think it's in line with a quote that NVIDIA's CEO said, "AI will not steal jobs, but people who know how to use AI will steal people's jobs." I think that's the framework that we should be looking and thinking about in this conversation. I know that we touched on this earlier already in terms of continuous improvement, but I'm wondering if you can elaborate a little bit more on what are some of the mechanisms that can be employed to ensure continuous improvement and accuracy of responses from ChatGPT, especially in the medical education setting where the implications can be life or death?

Dr. Victor R Lee (guest speaker):

There is a process in practice called fine-tuning where you could take the underlying large language model that ChatGPT uses or working with some of the existing chatbots to adjust and train it even further on very specialized texts. So that is one mechanism and one approach just to add or strengthen the likelihood that certain topics and words are appearing in frequency in certain ways when it responds. Regardless of that as being another process to enhance the technology itself, I firmly believe humans must always be in the loop, especially with high stakes decisions, and this needs to go beyond what do I think in this individual case for a patient, but also with larger groups of people.

What is represented in there is a product of what text it has worked from in the past, and the reality is that not everyone's voice exists on the internet in the same forms, in the same ways, in the same amounts, and that is reflected in the technology. So we need to be thoughtful and critical about what

this can do and what it does represent, and while we can fine tune it using additional text, again, it's based on what is available. And until we have a more equitable society with far better participation representation in all of these different spaces, we're going to be quite a ways off from being able to dismiss that and thinking about where are some of the more distal effects of this that may come indirectly.

Ruth Adewuya, MD (host):

One of the things that I just heard you say that resonated with me is in line with the idea of human-centered AI and the concept of how human input and collaboration continues to be at the center of the innovation. But also in line with the skepticism, my next question has to do with ethical concerns that you might have regarding using ChatGPT or similar generative AI model integration into medical curricula.

Dr. Victor R Lee (guest speaker):

Let me break that into two responses. First off, a number of analogies are coming out for how to best think about this kinds of technology. And with respect to this being human centered, what does that mean and how could we relate to or understand that? Does it just mean humans are placing checks on it? Does it mean that it's just built to be really easy for the things that humans have to engage with? The analogy that I see getting a lot of traction I think is a pretty decent one, is to think of AI as being more like the electric bike and less like the robot vacuum. The robot vacuum means that you aren't really doing the vacuuming anymore, you just give it to the technology to do. The e-bike is something that greatly empowers, but you're always in control, you are the one that pedals that lets you go faster, you steer, you choose the route on that.

So that's a way to think about how can we maintain human centeredness, so it's not about just trying to replace. We all can imagine and maybe have read some of the horror stories where robot vacuums mess up or cause more damage in the home than actually solving problems. Whereas e-bike, we can appreciate that we can still get exercise, that we can get to places we want to faster, that we can make decisions in the snap moments about safety, about how we want to have the biking experience to go. So that's a way to think about human centeredness.

Ruth Adewuya, MD (host):

I really like that. That crystallizes things.

Dr. Victor R Lee (guest speaker):

Credit due, that's an analogy that's circulating a bit in the field. The ethical concerns that we want to think about, I am hopeful still that the introspection that especially American society is taking part in continues and thinking about where there are biases that exist in many different forms. So while I am myself not a medical practitioner, I have seen a number of studies and reports about who's represented in medical textbooks, when you look at dermatology, what skin complexions are included in there, how do implicit biases based on perceptions of different groups lead to how we conceptualize who is in pain and how severe their pain is. And I think that there's a real risk here ethically of reproducing those by relying too heavily on something that's trained on these data and is not up-to-date, because it takes a long time to build and train these models, and you have to use stuff that's already existed in the past to have that sheer volume of it.

So I do think that with respect to medicine, aside from risks of just simply inaccuracies and out of date medical information, that it would serve to reproduce existing biases that have existed and that communities have known have existed. It also by virtue of not necessarily having everybody represented different sets of values and perspectives and experiences with medicine, the medical establishment may not necessarily be in there, and how would a patient actually interact, really it's a patient who is a conglomerate of the people who write actively on the internet, and that's a very particular population of people. And if you do work outside of the United States or in areas that have even more visually different cultural differences that we don't necessarily know or have that documented in writing. And so if we tend to be too reliant on the AI technology, we're at risk of being very naive or very myopic with respect to how we see things which are much more nuanced and complicated.

Ruth Adewuya, MD (host):

It sounds like there is something that everyone needs to be on the lookout for, be aware of, especially in medicine, that as we revise content or as we use ChatGPT for case-based learning for vignettes that are used in education, that we're aware that it does not introduce or reinforce existing biases. My last question for you is around faculty. I think that we're in this age where innovations are fast and furious in emerging. How would you approach faculty who are resistant to using ChatGPT and similar chatbots, many of whom believe that the platform's threaten academic integrity? What would you say and how would you have that conversation?

Dr. Victor R Lee (guest speaker):

First off, chatbots is just one facet of this new genre of generative AI technology. If we think just about ChatGPT and chatbot sorts of text-based technologies, there's a sort of inevitability here in that we can't really turn back this massive tsunami that is coming. And students are going to use this anyway, so could we look at this as the opportunity to use it in the most productive and beneficial ways for their learning and for them as future practitioners? So I would make an appeal on the basis of that. I would also gently question them about what is so important about the specific assessment forms, whether it's written essays or briefs, that make that the best or only way to know what students know or see what sorts of progress have been had and to empower them to think that this is actually a very neat time to be creative and exploratory.

The other thing is that there's fear about academic integrity, there's fear about cheating, it's actually an area that I'm doing a little bit of research on right now. And while we're still processing the data, there's very good moral compass we're seeing with respect to what people feel is and is not acceptable. And when there are things that seem like they're a bit slippery, there's larger factors to look into as to why is that even the case. We're not blank slates and simply sponges. It's a complicated process by which we learn and how we can best leverage the scholarship on learning, and we might be creating situations that are just causing more stress, affecting people's health and wellbeing, and ultimately just not letting them have the time or space to do the meaningful learning.

With respect to academic integrity, I would say it's not as dire, we're not seeing evidence that it's as dire as the headlines and the pundits are saying, even just from our current empirical work in the projects I'm part of, I think it's also a chance to revisit and ask why we do what we do and how can we all work together to make something really amazing happen for the benefit of future clinicians, for future patients, and for our own work as educators where we can actually find some ways to streamline things so we can really invest as much time as possible in the things that we love and that really need a direct human touch.

Ruth Adewuya, MD (host):

I think that's a great call to action for clinicians, for educators to stay curious and stay engaged in exploring how this can be another tool in our quiver. So Victor, thank you so much. This was very informative for clinicians and educators as they navigate what is ChatGPT and what are the implications for education.

Dr. Victor R Lee (guest speaker):

Glad to be here. Thanks for having me.

Ruth Adewuya, MD (host):

A couple of weeks after my conversation with Dr. Victor Lee about the impact of ChatGPT and education, I have the opportunity to chat with Dr. Nigam Shah. And I specifically wanted to chat with him about how ChatGPT might potentially change clinical medicine and healthcare delivery. Here's a short bio of Dr. Shah and what he had to say. Dr. Nigam Shah is a professor of medicine at Stanford University and the chief data scientist for Stanford Healthcare. He holds an MBBS from Baroda Medical College in India, a PhD from Penn State University, and completed his postdoctoral training at Stanford University. His research group analyzes multiple types of health data such as EHR, wearables, web blogs, and patient blogs to answer clinical questions, generate insights, and build predictive models for the learning health system. At Stanford Healthcare, he leads artificial intelligence and data science efforts to advance the scientific understanding of diseases, improve clinical medicine practice, and orchestrate healthcare delivery. Dr. Shah, thank you so much for chatting with me today.

Dr. Nigam Shah (guest speaker):

Oh, absolutely. This is a fun topic.

Ruth Adewuya, MD (host):

People have been wondering, speculating about the impact of ChatGPT across multiple industries, and so today hoping to unpack its impact in clinical medicine. Can you tell us what are some of the ways that ChatGPT is currently being utilized in clinical medicine?

Dr. Nigam Shah (guest speaker):

ChatGPT took the world by storm around November of 2022, and as far as I would say, it is not formally being used in clinical medicine. And I say formally because you bet all of the residents physicians have tried stuff. I've put in my own questions about my own care into ChatGPT just to see what it says, not really inform decision making, more for entertainment value. For our listeners, the thing that I really want to convey is that ChatGPT is one software application that runs in the web browser that can use one of a few large language models in its backend. It can use GPT-3.5, it can use GPT-4. It's important to make this distinction, because ChatGPT is like a user facing thing and then it has an engine inside. And if we ask the question how many such engines exist, there are over 130 or 150 large language models out there, and most people find that shocking.

Ruth Adewuya, MD (host):

Absolutely.

Dr. Nigam Shah (guest speaker):

The big ones are obviously from OpenAI and then there's from Google called Bard and Med-PaLM, there's one or two from China, and then there's a whole bunch of open source things now usually go under the name of some animal, like LLaMA, Vicuna, Alpaca, things like that.

Ruth Adewuya, MD (host):

I'm glad that you're clarifying for us what ChatGPT really is, and I definitely was surprised there are hundreds of them available. I'm curious to hear from you what were some of the biggest challenges? In thinking about the incorporation of ChatGPT in clinical medicine, I want to make sure that I heard you, you say it's not formally being used, but as people are experimenting, what are some of the biggest challenges that you are seeing?

Dr. Nigam Shah (guest speaker):

I would put the biggest two as we don't have a strategy to evaluate it, and then the second is we don't have a strategy to spot its errors and remedy them. So those I would say are the top two.

Ruth Adewuya, MD (host):

That's very interesting. I think those are major challenges to have to navigate, especially when you start thinking about clinical medicine and it's impact on patient health.

Dr. Nigam Shah (guest speaker):

Exactly.

Ruth Adewuya, MD (host):

And so it sounds to me there's a lot more we don't know for us to even consider its formal use in clinical medicine.

Dr. Nigam Shah (guest speaker):

So just like we surfaced this issue of ChatGPT the application and the engine inside the language model, a lot of people would be surprised that GPT, the language model, the first time it came around was 2018, 5 years ago, but those things were hard to use. And so I think the genius of ChatGPT was to make it accessible. And there's value to that, making things accessible so a lot of people can try and test, and that's why we're having this conversation. But the underlying tech is the same, it's based on something called Transformers that Google released open source in 2018. And so it's really important to have this distinction, because since November 2022 we're all trying this out, and the general version of the study that you might encounter is that, oh, I used ChatGPT for these five things and it looks good.

Let's use the analogy of driving. So driving is an activity, there's a human and a car usually, or let's just stick with consumer driving, and we send the human to the Department of Motor Vehicles and they take a multiple choice knowledge test. We send the car... Or not we, the company that makes it sends the car to the National Transportation Safety Board, and then they do some crash testing and rollover testing and all of that stuff. Then we put the car and the driver together and take them for a road test. We don't send the car to the DMV to take a multiple choice exam. See where this is going?

Ruth Adewuya, MD (host):

I do see where it's going.

Dr. Nigam Shah (guest speaker):

So what we're doing right now is we're taking the car, in this case GPT, sending it to the DMV and have it take the exam, the USMLE, or whatever, some board or what have you, and if it passes, we declare that thing is fit for medicine. It's equivalent to sending a car to the DMV, have it take the multiple choice test, and certify it as safe for driving. That is equally absurd.

Ruth Adewuya, MD (host):

That's a really helpful distinction in terms of where we really are in the process of this being a thing in clinical medicine. I'm wondering, as people are thinking about the different ways that it can be potentially used in medicine, are there some specific medical specialties where they forecast that it might have more utility?

Dr. Nigam Shah (guest speaker):

So let's take a step back. Instead of medical specialty, because they keep changing, today we have this many, tomorrow we might have a new specialty, let's focus on the task that these specialties have to do.

Ruth Adewuya, MD (host):

Excellent.

Dr. Nigam Shah (guest speaker):

And we can break down the tasks into summarization of information and generation of information. So for example, if you have a message from a patient asking something, you have to generate a paragraph to respond. That paragraph did not exist and you have to create it. But if you're in the hospital and you're doing handoff from day shift to night shift, there's the entire day's worth of events already recorded in the EHR and you have to summarize them for the night shift. Or when you're referring a patient from primary care to specialist, there's a whole course of history of present illness that has to be summarized for the specialist to appropriately accept the patient. So I like to make this distinction because language models are great at summarization, they tend to make fewer mistakes because they have guardrails in the form of the pros, the 1000, 2000 words that they have to then summarize into 500 words.

Whereas generation, all bets are off. The whole point of generation is to create things that did not exist. Language models create things or create the words based on the general patterns of what words go together. And so they will make stuff up, otherwise called hallucination or confabulation, and there's no guarantee that it'll be factually correct. And as we think of use cases in medicine, I would say summarization use cases are way more potential and in shorter time for us to be able to figure out is it working, is it useful, what can it do for me.

Ruth Adewuya, MD (host):

I guess there's this misunderstanding of what the tool is, its capabilities and its impact, because I think some of the conversations that I've had socially around this is concern about what does this mean for jobs, what does this mean for clinician roles, is this AI thing taking away jobs and roles and all of that. What's your response to that question?

Dr. Nigam Shah (guest speaker):

So naturally these things, particularly the user interface, gives it this magical veneer of these things are amazing, they can replace a human. I think those fears are overblown. I would go as far as to say that there will be certain job functions of people that will be replaced, and for good reason. And there's nice studies out of MIT department of economics analyzing a few white collar jobs, so to speak, to say, "What does the use of these tools do to the nature of work?" And two findings stand out. Finding number one, that it helps lower skilled workers a whole lot more than it helps high skilled workers, which makes sense. Imagine if you're a copy editor and if you're early in your career, a summarization tool would help you a whole lot more than if you're further along. And then the second thing that came out is that even for the experienced people and also for the junior people, it changed the nature of work.

If you're writing an article and typically whatever hours were spent on drafting and then on editing, the use of the tool shifted the hours, people spent less time drafting, but more time editing, fact checking and proofreading. Overall, across the board in many studies people are seeing 15 to 20% increase in productivity. So I think the impact of these technologies is real, no denying that, but I think as a profession, we have to not fall for the new shiny object, flash in the pan, whatever analogy you want to use, and ask the question what parts of our job this thing can do for us so that we can take drudgery out of the system.

Ruth Adewuya, MD (host):

That's a very powerful way to look at this tool because I think that when you start thinking about some of the functions of clinicians synthesizing that information and working to diagnose a medical condition, or even when you talk about personalized treatment recommendations, or when you think about can it assist in predicting disease progression and things like that, what I'm hearing you say is can it do all of these things, probably not. But what it can do is maybe assist in some portion and hopefully work alongside the clinician to help them do their job more effectively. Can the tool diagnose, probably the answer is no, but can the tool help the physician synthesize the information that they have on the patient so that the clinician can diagnose? I think the answer is yes. Is that what you're saying?

Dr. Nigam Shah (guest speaker):

Directionally, absolutely. Let me add a little bit of more color to that. Whenever AI comes up and job displacement comes up, the obvious answer, people go, "Well, we'll use these things to assist humans." But the nature of the assistance matters. If the assistance is such that I still have to proofread or check everything it does, it's not really offloading me. And so the way I think of it is that instead of having a physician in the loop, or human in the loop, as people like to talk about it, I like to think of it as what part of my job I can completely outsource to it. So coming back to diagnosis, let's say I'm going to see 10 patients today, some will be hard, some will be not so hard to diagnose, I don't want that thing assisting me in every 10 situations. What I want it to do is take the easy stuff completely off my plate.

So let's go to imaging as a easier to grok analogy, which Dr. Curt Langlotz, one of our radiology professors talks about, when he comes to the morning for his shift, there's 100, 120 x-rays from the night before that he has to go through. And typically he'll go through them in sequence. It doesn't take long and maybe he's done in three hours, two hours, what have you. But for the simple stuff, as in normal x-ray, can we just have an algorithm that just goes, okay, is it normal or not normal? If it's normal, we're 100% sure it's normal, don't even show it to the doctor. So then when Dr. Langlotz shows up, he only has to look at the 40 that are not normal. So we have to think about work differently. Now trying to add on AI in a sense that it's assisting him in every 140, it's probably going to increase his work.

Ruth Adewuya, MD (host):

Because now he's double checking. That's a very helpful framework to think about these things. And that lends to my next question around ethical considerations that we should be thinking of when in the example that you gave where Dr. Langlotz is not looking at the radiology, what are some of the concerns that you've heard around that concept? Because I can see implications for billing because usually they bill based off of the radiologist's time, and so it's like, "Are X-rays free now?"

Dr. Nigam Shah (guest speaker):

It's a great question. So today we think of these things as augmenting or assisting humans for a lot of reasons, billing being primarily one of them, because if it's autonomous, we can't bill in today's environment, but there's hope. Of the 500 or 600 FDA approved software as medical device, so we're not talking exactly language models, but AI in general, there's one that can work autonomously and bill $85 for checking somebody's retinal scan.

Ruth Adewuya, MD (host):

Really?

Dr. Nigam Shah (guest speaker):

Yes.

Ruth Adewuya, MD (host):

Wow.

Dr. Nigam Shah (guest speaker):

Built by Dr. Michael Abramoff at the University of Iowa. So there is precedent that we can create AI that is narrow, it knows how to do three things, and it knows that if it's not one of these three, I need to escalate and refer to a human. But we haven't organized our work around this way of partitioning things up. Back to language models, we could use language models to always write out post-discharge or post-visit instructions at the reading level and the language of the patient's choice. We already have post-visit instructions in a format, so it's grounded, we have a reference point or a framework, and it has to be summarized or translated. Chances of errors, small. Or going back to the x-ray example, if the only thing it's doing is is it normal or not normal, we can have that done by the algorithm.

We can then ask questions about billing, what are we going to do for this job role that the normal x-ray reading is now, say, a dollar, whatever, we can then test its safety, have it do the past thousand x-rays and see how often it makes a mistake, which then allows us to set a risk tolerance threshold. It then also allows us to test it for bias, and is it getting the assignment of normality wrong for people of a certain type or certain diseases or taken by a certain kind of machine? But if we define the role to be constrained, it lends itself to thinking about regulation, evaluation, billing, all of those things. Right now we're viewing these things as general replacements for humans, and they're not ready.

Ruth Adewuya, MD (host):

Does it feel a little bit like the Wild West in terms of what's happening in the AI space or in even with language models where everybody is just interested in what's the next new thing and so there's no strategy?

Dr. Nigam Shah (guest speaker):

Yeah, absolutely. So I think it takes the whole notion of the wild to a whole new level. And just to paint the comicality of the situation. So there's 180 or so professional medical societies, give or take, each and every one is probably having a position statement on how thou shall use GPT in your work, there are multiple engineering professional societies that have issued proclamations on how thou shall live life in the regime of GPT. Every journal has a review and multiple perspective and opinion pieces, almost all of which is based on anecdata. So it is in some sense a mad rush to be first. And the thing I would caution our listeners against that, it is no fun being first if you're wrong.

Ruth Adewuya, MD (host):

Is there work being done to create the strategy to build the frameworks, or do you see that as something that will happen in the near future?

Dr. Nigam Shah (guest speaker):

Well, it's already underway. Multiple societies are working on it. So the National Academy of Medicine just launched what they call AI code of conduct, which is broader than just ChatGPT or a language model. There are multiple health institutions talking to each other, trying to figure out... Imagine a triangle where one node, one vortex is building these language models, and ChatGPT being one of possible 100 of them. Another vortex being verifying their claimed benefits. And then the third being putting them, deploying them in health IT systems and asking did they deliver as advertised? So if you think of this triangle, there's a lot of activity in all of these three corners of this triangle, so to speak. In the media, we typically hear about the building of the models, and every 24 hours there's this company build this new large model, which is now free and open source and this and that.

On the verifying benefits side, I think that is where we as a medical profession need to focus and get away from these anecdata experiments where I took the pediatric board exams five questions and submitted them to ChatGPT and, look, it got four out of five correct. I was like, "I think 100 papers have been written and we can move on from that now." And then the one which is deploying them in production and then actually asking did we get the advertised benefit? That's not being talked about much actually. And so if we think of this triangle as a homunculus, which our medical listeners will follow, we're way overweighted on the building of the models and writing cartoonish papers. So we're climbing the slope of the hype cycle, and most people are familiar with the Gartner hype cycle, and I would say we're climbing up so fast that we might reach escape velocity and never come back down.

Ruth Adewuya, MD (host):

Thank you so much for chatting with me. This has been a very enlightening conversation.

Dr. Nigam Shah (guest speaker):

Well, I hope it helps our listeners, and happy to do this anytime.

Ruth Adewuya, MD (host):

Thanks for tuning in. This episode was brought to you by Stanford CME. To claim CME for listening to this episode, click on the claim CME link below or visit medcast.stanford.edu. Check back for new episodes by subscribing to Stanford Stanford Medcast wherever you listen to podcasts.