Adrian Hernandez:

Hey, this is Adrian Hernandez and welcome to the NIH Collaboratory Grand Rounds podcast. We're here to give you some extra time with our speaker and ask some of the tough and interesting questions you want to hear most. If you haven't already, we hope you'll watch the full grand rounds webinar recording to learn more. All of our grand rounds content can be found at rethinkingclinicaltrials.org. Thanks for joining.

Hi. Today we're here with Michael Pencina and Brian Anderson, who will be continuing our discussion from a recent grand rounds session on public-private partnerships in the trustworthy health AI ecosystem. I'm Adrian Hernandez, today's moderator for this podcast. So, Brian and Michael, hey, really thank you for sharing what you all are doing, leading some important efforts for the nation around trustworthy AI. First, what's the problem that you all are aiming to solve? Brian, you want to start first?

Brian Anderson:

One of the biggest issues that we're facing in CHAI is developing consensus definitions around what good responsible AI looks like in health. We have very high-level violent agreement around concepts like fairness and transparency, and what those mean. But what we don't have is the technical layer of detail, granularity that would inform an ML ops manager or a software developer. So, one of the things we've launched since we last talked is five working groups that are focused specifically on that, developing these best practice frameworks for generative AI or predictive AI, as well as developing the set of testing and evaluation matrix. So, once you have that definition, that technical definition of what good responsible AI looks like, how do you come to an agreement on what some of the specific metrics are? How do you measure bias in generative AI? How do you measure performance in an LLM note summarization or ambient listening device?

These are things we don't have consensus on, but we urgently need to develop the kinds of metrics and common agreement about what those look like. And so, we've launched these working groups. We started them three weeks ago and we're pretty excited about the path forward. And one of the things we announced recently is in October, we're going to have the first public draft available for everyone to have.

Adrian Hernandez:

Terrific. And Michael, you come at this from a different angle, a different experience. What are the problems that you see that you all are aiming to solve?

Michael Pencina:

The biggest concern and the biggest opportunity for us to contribute is bringing some order to chaos. We came back, Brian and I, and many others from CHAI convening during Stanford AI week and talked to a lot of health systems, as well as other stakeholders. And the message that I've been hearing is we need help to organize the wealth of tools that are being developed and thrown at us. And that goes for ambient voice recognition with DAX, Abridge, Suki and so on so forth. Sepsis algorithms, where you have individual health systems, you have startups, you have health IT technology developers. And health system users are asking questions, how do we know which ones to use or even in which areas these algorithms will be helpful. So, my dream solution is some kind of a registry and a marketplace of experiences that bring clarity to the chaos we have right now.

Adrian Hernandez:

Well, we'll circle back to that a little bit. One of the things that you all discussed is this concept of ABCDS. What is that aiming to do?

Michael Pencina:

ABCDS is our algorithmic governance at Duke. And I will say from my perspective, ABCDS was a big impetus to start the Coalition for Health AI thinking that algorithmic governance, what we have at Duke we want to share. And then like-minded partners who helped found CHAI were on board. The roots of ABCDS, the design of it go back to medical devices. And the process we put together follows what the FDA has been doing with evaluation of medical devices. When we were starting ABCDS as our governance, I was joking that we're playing mini FDA on the local level in our health system. But I think the exciting part of ABCDS is with the revolution of ChatGPT and large language models, we were ready in the sense that we had governance. We were not ready in the sense that it's constantly changing. We're adapting the processes because the new tools and technologies require different approaches, different metrics, different ways of thinking.

Adrian Hernandez:

Now, Brian, you not too long ago became CEO of this thing called CHAI. Tell us about CHAI. What is it? And why did you decide to take the leap to lead CHAI?

Brian Anderson:

When I was at Mitre, I think one of the wonderful things about Mitre is it affords you the opportunity to work at the intersection of public partnerships and private partnerships. And while I was there, to your point, we founded and we launched this public-private partnership. But at the time, it was a coalition of the willing. It was not a formal entity. And it started with eight of us, Michael and I, and six others. And it quickly grew to where it is now over 2,500 different organizations, the largest health AI coalition on the planet. To engage that many smart people that are excited to lean in, you can't do that with purely volunteer effort.

And so, I think recognizing the need to form this nonprofit and also appreciating that it needs the kind of leadership that I would hope enables organizations that are part of CHAI to trust CHAI. I saw an opportunity to help continue to further the mission. It's something obviously I believe in. I want our organizational members to find value in being part of this. And I'm really excited to lean in and help as much as I can. We're building an incredible team. Being in startup mode is new to me. I haven't ever launched a startup from the get-go, so I'm learning a lot. But really excited to have people like Michael on the team always willing to help and countless others that are there standing with us.

Adrian Hernandez:

Wow. Well, I'm just stunned to hear about the number of members already, because actually the last time I had looked at things, you had just crossed a 1,000. And I was like, oh my God, that's huge. But it looks like your doubling rate continues to grow higher and higher.

Brian Anderson:

Yeah, I hear AI is having a moment right now. We're certainly the shiny object in the room. I mean, part of the challenge is, is when you have this many organizations that are interested in it, is focusing on what it is that we believe is core to our mission. And so, it can't be that we do everything that everyone wants us to do. And so, part of it is really ruthless focus on what our priorities are, establishing a certification framework for quality assurance labs that are trustworthy, that can be used by the private

sector and potentially by the public sector at some point, but certainly initially by the private sector. And building those consensus definitions, what responsible AI looks like and what are the metrics that these quality assurance labs can use as part of their evaluation to bring greater transparency for patients, for providers, for health systems, for society writ large.

Michael Pencina:

What I will add is what I'm particularly excited is not only the number of members and participants of CHAI, but it's also the increasing diversity of voices. We started originally with primarily health systems, academic medical centers. And now it has grown with industry and it's the very big guys, but also a lot of startups, smaller companies. As well as health systems, the academic, the affluent as well, the ones that are serving the underprivileged constituents.

Adrian Hernandez:

So, you've grown tremendously to now over 2,500 members. How do you ensure the different voices have a seat at the table large and small?

Brian Anderson:

It's a great question. It's certainly something that's really important to us. When we started CHAI, I think one of the founding principles was that this cannot be something that is done just by academics. That this cannot be something that is done just by big tech. It cannot be something that is done just by regulators, that it has to be bringing those groups together. And so, part of the structural governance and organization that is core to CHAI, starts with our board of directors. If you'll notice, the big tech "representative," is Dr. Eric Horvitz, who sits on the president's council of advisors for science and technology, and is also the chief scientific officer for Microsoft. But he's one seat. There is also another seat for startups. An equal voting chair, just alongside big tech. Similarly, we have representation for patient community advocates. We have representation for health systems. And so, the hope is starting at the top with the organizational structure and then going down into our working groups, it's not going to be just big tech or just AMCs.

Each of these working groups intentionally was structured to be large. It's somewhat challenging, but intentionally wanted to make sure that we had a diversity of voices. So, these working groups are between 20 and 40 members large. And if you think about it, if you look at the big tech players, so Microsoft, Google, Amazon, maybe one or two others, those are only three or four or five seats. The vast majority of the other seats are going to be occupied by startup companies, by mid-size companies. And each one of those seats has an equal vote. So, it's actually, when you think about it, almost the other way around in that big tech does not have the equal representation that the larger ecosystem has because we're ensuring that the small players have a voice that is equal to the big tech.

And part of the reason why we did that is we did not want to create the kinds of best practice frameworks or T&E for testing and evaluation frameworks that would make it a challenge for a startup to adhere to those frameworks. And that could potentially, further down, enable the kinds of regulatory capture with regulatory frameworks that are so onerous on startups. And so, we're really focusing on ensuring that we have the diversity of voices, that we're including the little guy, both startups and small health systems, so that the work that we're developing, these consensus technical definitions can be easily adopted by organizations that have less resources.

Adrian Hernandez:

That's terrific to hear. So, I'm glad that the world is going to be inclusive here. Take us a little bit around how you all are thinking about validation of AI algorithms. It's not necessarily always an easy task, but there are obviously different ways to set things up. But what are the key issues?

Michael Pencina:

I think it's a very, very interesting space that's emerging in front of our eyes. And I'd like to think about the pre-deployment stage and post-deployment stage of algorithmic evaluation. And I've heard some proposing the notion that, okay, there are algorithms and there are tools, and you health systems or users go and do your own validations. And I think there is value in the local evaluation. I actually published a paper with Nigam Shah from Stanford and others calling for recurrent local validation as a new standard for algorithms. So, we absolutely need that. We also need to focus on outcomes that measure patients' mortality, morbidity experience in that. And I think academic medical centers and health systems can do a very good job on the monitoring and the experience side.

But that cannot be all. We cannot just leave it to the users to evaluate or validate what's being proposed. The developers need to be a part of it. And it needs to be work that really is joined, because our experience for medical devices, right? Take a stent as an example. It's not that a manufacturer produces a stent and tells a health system, "Well, if you want to use it, you do your validation and decide if it's safe." That's not the process that we're used to. So, I think CHAI gives the opportunity of different constituencies on the developer and the user side to come together and define what the process looks like. What are the standards? We'll be releasing our assurance standards guide in a couple of weeks from CHAI on what needs to be part of the development process, and then the validation, and then the monitoring that happens on the level of the users, health systems, and a feedback loop to the developers. So, that's my dream ecosystem and all of that with registration, keeping track of the tools that are being produced and deployed.

Brian Anderson:

Recently, since we last spoke, a couple of, I think, really important bookmarks have happened. So, the first was the Office of Civil Rights within HHS announced a update to section 1557 of the ACA. And they clarified that algorithms, if found to be discriminatory or have unjustified bias in them, that the provider organizations' providers are liable, can be held liable for lawsuits and action by the Office of Civil Rights. That's a real legal exposure. What I am hearing, what I'm personally hearing from a number of the health systems within CHAI is that the AI governance committees being informed by the general counsel's office at these health systems are now demanding from vendors to have external validation of those models' performance. That it is not good enough for the model vendor to share their own internal testing, but that we need independent validation. And this is the same principle that was echoed in 2021 with the White House's AI proposed Bill of Rights and need for independent validation.

We see it in every other sector of high consequence in society. Automobile driving, nuclear reactors, consumer electrical devices in our homes. These things are independently tested and validated. And we need that kind of external validation. In health AI, we don't have that. We're seeing health systems already demanding it before they make procurement decisions. To say that a health system can only have the kind of validation that is local, which implies that they would make the procurement, that they would deploy it even in a pilot setting, that they would go through the expensive configurations needed to actually do that kind of testing, I think is narrow-minded and shortsighted.

What we need is we need both. We need the fantastic work that Epic is doing with Seismometer, creating a open-source community that is going to start sharing the kinds of tooling to do the recurrent local validation that Michael's talking about. We absolutely want that. We need that and we need to do

that tooling in a way that enables lower-resourced health systems to be able to actually do that as well, to do the governance and monitoring that Michael's talking about. But that can't be sufficient. For organizations that don't have the kinds of resources, that don't have the kinds of tech talent, and for organizations that want to be informed before they make a procurement decision, you need external independent validation as well. So, we need both of them.

Adrian Hernandez:

That's good to hear because when you describe some of this, and especially in parallels to other industries that we take for granted that they have safe technology, we certainly need to do that for health. Related to that, there's an ocean of possibilities around AI and health. You all have started with healthcare systems. But what else is out there? How are you forecasting the future in terms of incorporating that really incredible sea of possibilities that could affect health or improve health?

Michael Pencina:

Well, I think the reason we're at the table is the opportunity. And we talk about the valuation, governance, best practices, but it's really at the service of innovation. We want the field to flourish. And as we know, our healthcare enterprise nationally is at risk. I think the pandemic has brought forth a lot of issues, the issues of cost, the issues of burden, the issues of access to care, and I think AI has a tremendous opportunity. What I've been seeing is a shift from development of clinical tools to development of operational tools for AI that reduce clinician, both doctors and nurses' burden, opportunities to use AI for better claim processing. But there are risks. We don't want AI to be used to deny claims. We want AI to identify the procedures that patients truly need and work for the benefit of the patient. So, I think that's the healthcare that I'm hoping for and that will be here for us with the help of AI.

Brian Anderson:

For me, Adrian, there's this term that some of us in CHAI have used in the nascent space of Gen AI and ChatGPT, and Gemini, and all these models that many of us are playing with, the concept of privileged access. So many of us go online, use these tools, the well-educated, well-informed and health systems that are well-resourced. What I'm excited and we're beginning to see some of the fruits of is moving from this concept of privileged access to just access. Meaning that it becomes more mainstream, that it becomes more accessible, more usable, more intuitively understood, and navigable by more than just people like three of us on this podcast. And I think that's really, I think, going to be transformational for our society.

Particularly when we think about individuals sitting at home who don't have the ability to go see a doctor right away. Maybe because they just don't have a doctor in their community and it takes an hour or two to drive to a community that might have a doctor. Those individuals being able to go online and navigate, and get responsibly-developed AI-informed recommendations, or guidance, or education that would then empower them to take whatever the next step is, that's really exciting to envision and see how that will transform not just the healthcare delivery system, but just health and how we think about health in our societies writ large. The thing that concerns me about it is with greater access, you have greater demand. And we are not creating enough nurses and doctors fast enough. And so, how we bridge that gap with greater access and meeting it with the supply side is going to be a real challenge. I think we need to do things like what Michael just talked about, the efficiencies on the backend, operationally. I think that's where we really need to focus and lean in more.

Michael Pencina:

All right. Great. Well, last question, and this may be a hard one. Normally I ask what are your predictions from five years from now. But this world is moving so fast in this space, I'm going to ask you, what's your predictions here for CHAI in three years and trustworthy AI?

Brian Anderson:

The more and more that we talk to and spend time with the developers of some of the frontier models that are out there, the more I'm excited about not three years, but what the next two years are. I'll make somewhat maybe of a bold prediction. I think we're going to have real tangible AGI within the next two years. I think that one of the things that that's going to mean for CHAI is how we think about responsible AI and certification of intelligence that is indistinguishable almost from how doctors describe and think about our intelligence. And so, when we think about it from that perspective, the evaluation of AGI-based models, I think, is going to start looking more and more like how we think as clinicians certifying clinical competencies in medicine. We'll see, but that's a prediction for you.

Michael Pencina:

I would say my hope for CHAI is bringing order to chaos. So, imagining CHAI as this marketplace that through the local registries that we're aggregating on the CHAI level, you will be able to go and look at any given tool and read about the key information about it, the experiences of different health systems, and how it fits for potential use. And then having, with our assurance guide and certification, people getting comfort and saying, oh, yes, CHAI or one of the CHAI assurance labs looked at this, and this has been validated through the CHAI process.

Adrian Hernandez:

Terrific. Brian, Michael, thanks for sharing your insights and a bit behind the scenes of what you all are doing in terms of developing trustworthy health AI ecosystems. So, it's been terrific. And I hope everyone enjoyed this podcast. Please join us for our next podcast as we continue to highlight fascinating changes in the research and healthcare world.

Thanks for joining today's NIH Collaboratory Grand Rounds podcast. Let us know what you think by rating this interview on our website. And we hope to see you again on our next grand rounds Fridays at 1:00 PM Eastern Time.