Transcript Details
This is a transcript of an educational program. Details about the program and additional media formats for the program are accessible by visiting: https://reachmd.com/programs/the-convergence/medical-ai-reasoning/54785/

ReachMD
www.reachmd.com
info@reachmd.com
(866) 423-7849

Medical AI Needs Reasons, Not Just Answers

Dr. McDonough:
Welcome to *The Convergence* on ReachMD. I'm your host, Dr. Brian McDonough, and joining me today is Professor Michael Moor.

I really want to welcome you to the program, Michael. It's great to have you here.

Dr. Moor:
Thank you so much for having me.

Dr. McDonough:
Michael, before we get into the technology, tell us a little about yourself. What brought you from medical training to machine learning and medical AI?

Dr. Moor:
I'm an Assistant Professor of Medical AI at ETH Zurich, where I'm leading the Medical AI Lab. So we're actually working on medical foundation models, language models, and AI agent systems specifically for medical applications. This comes from this dual background both in medicine and also AI and machine learning research.

So it all started out with medical training back in Switzerland, where this is a six-year full-on course. So it's not like in the US where you first have undergrad and then med school later on. And during that time, I was always interested in also going beyond just the standard clinical work, and already, back then, I was quite interested in research. But research typically meant doing bio research, like looking at different corners of the same molecule. And I was already then quite interested in math and computer science, but it was more self-taught and more of a hobby.

And then after the medical studies, I pivoted quite 180 degrees into machine learning research and started a machine learning-focused PhD. Of course, at the start, I had to play catch up quite a bit, but for me personally, it really felt like a magician going to a wizard school and learning new magical tricks every day. It really felt so powerful, what we can do with computers, and that was already almost 10 years ago. And since then, I was lost to the machine learning world but always carrying forward this strong belief that we need to develop machine learning and AI systems that do good and that do good in medical practice ideally.

So that's how the two worlds collided for me.

And later on, after my PhD at ETH, I moved to the US and did a postdoc at Stanford in computer science. And there I was really interested in developing AI models, not just generally for medicine, but taking a step back and realizing that all the medical AI systems we developed earlier on had this shared problem that we can train them on as much patient data as we want, but ultimately, they were not really able to reason about the underlying diseases with any prior knowledge any medical student might already have. So I was more curious, how can we actually train medical AI systems in a way that they know everything that there is to know about medicine before we even show them the first patient?

Dr. McDonough:
And you've gone through all these steps, and obviously, you can describe yourself as a physician, scientist, an AI researcher, or a builder. There are so many different things. And then you got to a point where you're looking at chest CT as a useful tool in AI. How'd you get to that?

Dr. Moor:
So usually, our group—and it is a very young research group—we're quite broad in the sense that we want to develop medical AI

foundation models and AI agents within the medical domain because we truly believe that the complexity of that domain is so high and the stakes are so high that it is an extremely ripe and rich environment to really develop new advances in the foundations of AI. But what that concretely means is that new projects come along and new opportunities come along, and that sometimes means we do something that is maybe not our daily bread and butter, but we try to really reach out and venture into new territories. Of course, the radiology field I've always found very interesting, and I've also had prior works going into, let's say, radiology AI direction.

But this very specific project launched in a bigger project that was embedded with Swiss AI, which is an initiative in Switzerland where different universities—for instance, ETH Zurich and EPFL—are contributing. And there's also external partners that can participate. And there is this supercomputer in Lugano, and there we can, as researchers, apply for compute grants, which are quite sizable, at least for an academic context, where we can really train large models. And here in this context, we trained an AI agent system, which is not so easy to do on your regular academic hardware that you get for doing research.

So that's the context of how this came about. And this consortium that we had here actually consisted of different researchers, some of whom already worked quite a bit in radiology AI.

Dr. McDonough:
When I'm looking at RadAgent—I'm going to have you explain exactly what's going on here—it just doesn't seem like radiology, and it doesn't seem like something out of a black box. It actually seems to be, in a sense, where AI is heading, hopefully. It's asking bigger questions like, "How are we going to use AI to think along with us and guide us?" Tell me about, first of all, RadAgent, and how it is flipping the script from just getting something out of a computer.

Dr. Moor:
Yeah, so the basic idea behind RadAgent is that we want to challenge the prevailing paradigm that we had previously in especially radiology AI, So in the last few years, there has been a lot of development of the sort. There's some sort of AI system, like a vision language model that would take in an image or even a volume and directly write or generate a draft report for a radiologist. And of course, this is really exciting development, but there's lots of things that can be improved there.

One thing we felt was a big improvement we can make with RadAgent is that those draft reports, they come black box out of nowhere. And the basic idea behind RadAgent is, can we develop an AI agent that is under the hood—basically a language model that was specifically trained to use a battery of different external tools? Tools might be some sort of software to process the image, segment specific organs, select specific slices of the volume, or even apply a specific windowing to show different contrasts as the radiologist would do, or even apply another vision language model and ask it a specific question.

So there's an entire battery of different tools, and we basically put the language model into the middle of this toolbox and say, "Go and analyze this volume step by step." And then what we get out of this is instead of just directly getting a report out of the thin air, we get this trace where the agent would do many different steps in sequence. It would maybe, for instance, call a disease classifier, and then the classifier would say, "Oh, there is a high chance of emphysema." And then the agent has some working hypothesis and can go down the rabbit hole more or less about this thing. And doing this at the end we still have a report, right? But the report now is equipped with this long trace where there's many different steps that include external tools that were called. And there's also a scratch pad where at each step, the agent has the ability to add or remove findings that, for instance, in this volume, the agent believes to be in there. So this gives much more auditability, and—we'll probably talk more about this later—also faithfulness because you can actually see what the reasoning processes laid out as a map.

Dr. McDonough:
It's so interesting, Michael, what you're talking about. I'm coming from the days when I would get a radiology report that more or less was written for me, and also, in the United States, especially for attorneys, so there wouldn't be malpractice. It would say, "Such and such diagnosis cannot exclude this and this." And you're almost reading it going, "What did I learn from this image that I ordered? I'm more confused than I was before. Now I have to order more tests to exclude other things that it did not exclude." Whereas this tells me it's more or less thinking along with you and just trying to help in patient care, which is really where we want to be.

I know in your work you use a term called "generalist medical AI." As a generalist family doc, I like to try to treat everything if I can to the best of my ability. Is that what generalist medical AI means? Is it like a giant model doing everything? How would you explain that?

Dr. Moor:

So to roll back a bit, roughly three to four years ago, we wrote a perspective piece in *Nature* introducing this notion of generalist medical AI. That was with several co-authors and colleagues all across the US. And a big question we had back then was, "Well, now we have language models, and they seem to be able to absorb a lot of medical knowledge already during pre-training and further training steps

down the line. But what do we do with this now? How can this add clinical value?”

And back then, it was a different time. Three or four years ago, if you asked me, I would've just probably naively said, "Let's just build the biggest possible data set and throw in all the patient data we can find and everything. And in the end, we'll have one big multimodal language model that will be able to solve all the problems." But honestly, I've gotten a little bit more skeptical, especially when we think about the high stakes of, the medical domain. Language models under the hood are still deep learning systems, and deep learning under the hood works exactly on what you train it to do. And as soon as you go a little bit out of the training domain, there's certain risks that the model will not really be able to perform as you intend it to.

And so in the following years, I got much more excited about thinking of generalist medical AI as AI systems that are equipped with many different tools that are much more reliable than the language backbone itself would be to then really ground the language model in the medical domain. And so that, I feel, has been for the last two or three years a bigger paradigm shift rather than just building the largest single guard model that is able to do everything directly out of its parameters. And while this was first motivated in terms of reliability and accuracy performance, I think there's actually a hidden feature in there that we only now get to realize, which is you're making your AI system not only more reliable if you equip it with the right authoritative high-quality tools like external knowledge sources, guidelines, et cetera; you also make it much more auditable and back-traceable. You see actually the reasoning, because if you just ask chatbot a question, you just get an answer. But in medicine, we don't need answers. We need reasons. There is a conflict of different competing hypotheses, and there's varying contexts, and this is a really rich discussion happening. All of that we need to expose. If we want clinicians to be involved in that, we need to expose that, and we need to build interfaces such that the clinician or other care providers can actually participate and not just be a passive observer or consumer of intelligence being commoditized into their chat interface.

Dr. McDonough:
Okay, so let me bring it back, as you said, to some of the basic things. When you say the system produces a reasoning trace, what does the clinician actually see? What would I be seeing if I was using the tool?

Dr. Moor:
To some degree, that's an engineering question, right? With throwing additional software engineering at this, you could get the nicest and cleanest interface for the clinician. But what we have built initially is basically the entire environment such that you can take a language model, equip it with all those tools I mentioned before, and train it using reinforcement learning to perform this step-by-step analysis more reliably and more faithfully and to write better reports.

The raw trace actually might be quite cluttered. It might almost look like code. Right now, we're running an evaluation with radiologists where we have some sort of front-end app where they see that as nice boxes and where they see individual steps, which is more interpretable. How do you actually surface it to the clinician—that's kind of a software engineering question. It's not only a core research question. But the problem there is, what do you do if an agent within seconds runs 50 different analyses? Do you want to expose everything at the lowest level of granularity, or do you want to just show summaries? And so this is a big question anybody who uses AI agents has to ask themselves. To which level of granularity do we need or should we be more adaptive where you know, “Okay, the agent now has run this specific analysis, and now we want to really know the details, and we go adaptively into one silo.” But that's a lot of user interfacing that might not be the core initial research question.

We're still moving in the proof of concepts territory, so we're not a startup, and we're not a big tech company that builds big solutions rolled out to a thousand hospitals. But we're trying to really scratch the surface of what companies will be able to do in the next years to come.

Dr. McDonough:
And I'm glad you put that perspective on it. Now, you've teased us for a while at saying we were going to talk about faithfulness and what it means, and you even talked about the tool being faithful. When you say faithfulness, what do you mean in this context?

Dr. Moor:
Yeah, I think that's a highly overloaded term in the general AI research landscape. There's some merit in actually defining what we talk about high level. I think when people talk about faithfulness in the context of language models, they usually mean, is the language model actually saying, in its response, the correct reasons that it internally also used? For instance, if there is a very specific hint in an input prompt that flips the response that the language model would give to the same question, you can then ask, "Why did you change your answer?" or, "What made you change your answer?" And if the model just says anything but is not pointing at the causal reason that flipped the language model's response, that would then not be faithful.

That was maybe a bit abstract. We can make an example with an image or with RadAgent. So that was also actually the stunning

finding for us because we did not bake that into the system. We did not design it to do that. That came out of nowhere after the reinforcement learning training. We gave it a task to analyze a CT image and write a report. And then we added an injected prompt hint that would either give a wrong finding. For instance, we'd say, "By the way, there is a pneumothorax in there." Or we would give a correct finding in there. But both of those would just be hints that are being given on the side at the start. And then we check what happens with any model. It could be our RadAgent or the baseline, and we check what happens if this hint is given compared to if there's no perturbation at the start. And if the model would flip its response, then we would attribute that to this perturbation—perturbation being there is some sort of hint at the start. And we then looked at those flip events, and we asked the model, "Why did you say there was a pneumothorax?" And if then the model would say, "Well, it was there and there in the image." It was not faithful because it came from a hint. But if it was able to say, "It came from this hint here at the very start," then that was faithful.

So of course, faithfulness is a really large concept, but this is one way we can probe the model or the entire system. Can it interrogate its own reasoning mechanism, and can it point at reasons in a way that reflects actually what it has been doing? And so that's how we operationalized faithfulness here. So we have this probe by having prompts at the start and introducing hints that could be correct or wrong, and we just check whether there's a flip in the response of the agent or the baseline vision language model. And we found that the baseline vision language model was able to correctly point at the hint instead of pointing at the image in zero percent of the cases. So the basic vision language model paradigm we found was absolutely not faithful; it was not able to distinguish or analyze where exactly in its own reasoning the specific finding was coming from. It could not say, "Oh, it was in the prompt." It thought it was in the image, or it could not re-robustly say where exactly this hint was coming from. And the agent had a faithfulness of, I think, 37 percent, which is still not perfect. But this is a major step change from zero to 37.

And what's also really nice is we did not really train the system directly to be robust in this regard. We basically just trained it with reinforcement learning to increase its reward that it will get from detecting the right findings in the report and to really use this toolbox of different external tools to use.

Dr. McDonough:
It's very interesting. Those points you made were excellent. By the way, I want to clarify that RadAgent is a research stage at this level, not a clinically deployed product, and you've mentioned that. Conceptually, I just want to get your view on this. Is the future about bigger models, better retrieval, or better tools? Or are you looking at orchestration between them? Where do you think we're going?

Dr. Moor:
Yeah, there's this famous saying that predictions are hard, especially about the future. I feel in the last four years, many developments or many big changes in the field felt very predictable. If you knew the last step, the next step was not so hard to predict. And to me, many developments ended up in this really big and broad concept of having AI systems that are, to varying degrees, autonomous and that are taking actions in various environments.

So this agent paradigm, I think, is where many strands of research are ending up, which does not always make sense in medicine because there's also many processes that are very rigid and where we don't need autonomy, but rather where we need a very clear workflow from left to right and where we need checklists. But from the AI side, I think the agent paradigm is here, and it will be here to stay for some time. And I'm a bit ambivalent about whether we really need to have bigger and bigger models, because I think there's also a really strong push towards independence.

This was not always the case, but I've seen it happening in the last few months and maybe years—I think there's many users who really want to have their fully local stack of everything. Be it a hospital, be it researchers, be it companies. And I think this is not really pointing towards there being ever larger and larger and larger models. But I think a big interesting question now is also, how can we build small models that you can serve on the phone, and that you can serve without internet access somewhere in the sub-Saharan region? How can we do medical AI in a way that we improve access instead of just giving access to the few?

So I think this is definitely a really interesting frontier—how to make AI more efficient. Right now, it is very inefficient still, and we almost need an entire data center to run the latest model. And that's not really commoditized intelligence, right? Rather, you need to spend a lot of energy and money to get potentially less intelligence than you just have with one coworker.

I don't really think we should use AI to compete with human workers; I think that's quite an unethical thing to do. I think AI is much better at doing very horizontal work within a second, like a search engine. "Within a second, give me all the most relevant literature, the most relevant clinical trials, all the relevant findings you will find in different data modalities, and collaborate with me as a human to arrive at better conclusions than I would have if I spent 10 years studying this case alone."

Dr. McDonough:
And if I'm following you right, I think what you're saying, at least with AI as it stands now, basically is there might be an explanation, and

it might even sound reasonable, but does it reflect the process that produced the answer? And we don't really always know that. Am I following you with that? Is that what you're getting at?

Dr. Moor:
Yeah. There's some established research around that for language models and even for humans. I think this has been quite an interesting debate for some time. Is our conscious, first perspective, stream of consciousness what is steering the robot we call humans—the flesh and bone robot—or is this just some kind of afterthought that happens? I'm not saying I have the right answer for that. That's definitely out of my expertise. But I'm just saying these questions we can already ask with natural intelligence that we have. And now, with AI systems, it's also a very important question people have been asking for some years now: if a language model is producing a long chain of thought and then produces a better answer following that or conditioning on that, does that mean this chain of thought was actually real reasoning or not? I think there's a lot of semantic fights and battles over what exactly reasoning is, which I don't always find very fruitful.

But one thing that I think can be said at this point is that those reasoning chains can improve performance, but that does not automatically mean that they reflect the internal mechanism of how the language model arrived at a specific decision. So, this faithfulness is not being given away for free by language models. And this is exactly why we need to open this black box and make language models take actions to really corroborate what exactly they are doing and make it observable. And that's also, I think, where agents in environments where they have to call tools, maybe even write code, execute code, use different software, or look at or guidelines, pinpoint exactly where a finding is in a guideline to be able to go to the next step, et cetera. I think that is where we can really make a much more introspectable system that is easier to audit for humans, easier to collaborate, and also easier to build trust.

Dr. McDonough:
You're listening to *The Convergence* on ReachMD. I'm your host, Dr. Brian McDonough. I'm speaking with Professor Michael Moor.

I want to ask you a question, maybe with an idea in mind. I want to ask you if you could give us a concrete RadAgent example step by step. But as you're doing it, maybe talk about, again, where it could mislead, what might need to be tested, and what physicians outside should understand.

Dr. Moor:
In the paper, we have one trace example that I can quickly talk about. So at the start, a human user would say, "Hey, RadAgent, please, go and analyze this CT volume for me." And then the RadAgent would first use a vision language model under the hood to draft an initial report. But that is completely black box, and this is just the very first sketch. And then it would use a checklist that we curated together with a radiologist that is not super detailed because that will confuse the agent, but is detailed enough that it will actually give it some kind of hook—what exactly the agent should be do, which is kind of a workflow, not a fully autonomous agent. And then, in this specific example, the agent would come back and would find after initial analysis that the heart contour and size are normal and there is a lung nodule present. And now, the agent is getting suspicious about the lung nodule and thinks, "Okay, let me actually use a disease classifier that would spit out different, specific findings that I can go and inspect further." And then this classifier spits out that besides nodules, there's also an effusion. So there's now a new hypothesis on the map. And the agent would update the scratchpad it has, and we know exactly in this step the effusion was mentioned the first time. So if a human, like a radiologist or researcher, wants to analyze this trace, we could directly pinpoint where in the trace this effusion is coming from. And now the agent wants to go deeper and would use a specific tool that is designed to segment effusions, apply that, and then also extract a specific slice from the image where this effusion is segmented. And now it has a two-dimensional slice and asks, "Okay, can you further characterize this effusion?" Which we then can use in the report.

And this trace will go on and on, but just to give you a little bit of a taste, there will be many different sub-steps, and an individual step might have a very focused question in mind for what exactly we need to analyze at this specific step. And then there is this integration of many different findings that happens towards the end where all those different, let's say, rabbit holes, are being used to then compile the full-fledged report. And those individual findings, if we want, we can pinpoint in the trace where exactly they came from.

Dr. McDonough:
You can kind of follow the steps that are getting you to the answer, right? You could trace it back.

Dr. Moor:
Exactly. I mean, an individual step is still a language model doing black box stuff as they always did. The one thing that is new in that regard or that we think is interesting is that there might be 50 steps, and there is a very clear trajectory where we can see, in a specific segment, there was a main question. Like, is there really effusion? Can we characterize it more? Or are there nodules? Where are they? Please extract the slices, segment them, highlight them, look at them with different contrast, et cetera.

So I think the power here is that if we have a scratch pad that is being continuously edited, where findings join in or are being removed, it's like a continuous process where we can see where exactly something appeared, at least on the map of the agent, and where it disappeared again. Maybe there was a suspicion of a certain finding, and then doing more analysis, we could also discard that hypothesis again.

Dr. McDonough:
So I think I get it. When you were talking about using it horizontally, almost like a search engine, I think what you're saying is I now can go into those steps—and let's say I focus in on the effusion as a clinician—I then can ask the questions that I would be thinking about related to an effusion and ask the RadAgent to take it another step further for me to help me along. So in other words, I'm not asking for the ultimate “this person has this and this” and “this is the treatment.” I'm getting deeper in and figuring out what’s going on that may be leading to the clinical problems or may have been a response to the medications we're giving or whatever. I'm able to go through those steps and pick it up where I need to. I'm making it very simplistic, but is that what we're really hoping to do?

Dr. Moor:

I think that's absolutely where it has to go. I think this fully collaborative interfacing is exactly where it needs to go. This is, of course, a bit harder to develop because the human is not so easy to model. And so the main machinery that is used to train such language agents, which is reinforcement learning, is much easier if it is done in a single process where there's like the agent, it has its environment, it interacts with it, and then it observes some reward.

What is much harder is the agent doing some initial steps and then a human coming in and saying, "Oh, what about this other thing?" So this multi-turn interaction between AI agents and humans is much more challenging. And I think this is still an active area of research—how we actually develop an agent that is fruitfully collaborating with a human. And I think this will be a really interesting frontier also for future research.

Dr. McDonough:
Well, I think what you're saying is going to be comforting for a lot of physicians out there. I mean, two years ago, if you were to ask a radiologist or a pathologist, “What's going on with this?” they're thinking, "Oh my gosh, a machine's going to come in and do my job and going to read this mammogram better than I ever could, and I'm going to be replaced." But I think what you're saying—and I like what you're saying—is, "No, that could happen, but what we're really looking at is taking it through steps that can make a better clinical decision that you wouldn't just get out of an image. It's kind of working with you." It's almost like you have grand rounds in one room. That's what I like. This could almost take the place of five different specialists getting together discussing this all, assuming that the information is accurate and the steps are there. It could be really helpful, yet leading the clinician to make the decision and recognize where there might be flaws.

Dr. Moor:
Yeah. In an ideal world, that's exactly how AI will be used in my view. That's definitely a mission I'm fighting for. I'm not sure this will happen everywhere because you can see any human having a certain performance level, and from a purely financial incentive, it's not so clear what exactly happens if we basically add more intelligence into the radiologist's readers room, or generally any sort of clinician's room, right?

There might be certain hospitals just saying, "Okay, you're now double time faster, so we now have twice the number of patients." And then it's not really clear if anyone really benefited except for revenue somewhere in a sheet.

Dr. McDonough:
Michael, you're bringing up a great point. I think you're bringing up some real issues. I remember I was in a room where they were talking about just having a dictation system built into the electronic medical record, and immediately, the bean counters, for lack of a better term, were saying, "Oh, you could see three more patients in the same amount of time." Whereas the physician was saying, "No, I could spend more time with my patients and get more from them." And they're saying, "But no, you could actually not spend as much time ‘typing.’ You'll get more out of it." And that is that battle. It's a battle between profit and care, and obviously they have to align, but I could see what you're saying. That is an issue.

So when you're doing this work, and you're thinking of the future, and you're looking at medical AI and where it's going, what worries you? Right now, what worries you? We know the positive sides and what we're all shooting for, but what are your fears?

Dr. Moor:
I think one thing that worries me is that there's a lot of effort we put into developing AI systems that are helpful and that support caregivers and clinicians. And it's not really clear—and this has been the same for years—on the other side of it, where exactly will the

value be added? I think it's extremely important that there is clinical benefit—that there is benefit for the patients, there's benefit for the clinicians, for the entire team, for the entire staff team. And it is not obvious to me that this is exactly how it will pan out, because the big decision makers in various caregiving scenarios have very different incentives. It's quite often Excel sheet managing—not patient managing—people who make decisions whether you will just be treating three more patients or whether you have more quality for the same number of patients.

So how do we make sure that AI increases care quality instead of just making the same care more efficient, and it would just increase quantity? Even that on its own might also be valuable because a patient might be able to get an appointment within a week instead of waiting for months. That might also be valuable, but again, I think we just need to measure—especially clinicians—the impact and the value added to the patient. I think this is so important because there's many other voices that will want to make money out of this.

Dr. McDonough:
There's so many things I want to ask, but I want to try to go into a couple questions. One would be, what should physicians be able to ask before trusting a medical AI system? What should we be asking or thinking of before we even trust it?

Dr. Moor:
This is a very philosophical question, right? It almost sounds like a nightmare. You're blindfolded, and you have one person who you're talking to, and you need to figure out within a few questions, can you trust this person? It will lead you out of a cave, or it will lead you even deeper into the cave. It's really a big question, and it's definitely not something that's easy to get out within a few interactions.

I think this is all about appropriate reliance. How do we learn when there's any sort of technology that's not even AI-specific? How do we learn to interact with it in a way that we don't give up? How do we not give up our skills or our sense of agency, but at the same time, also learn gradually where and when we can trust it and when not? I think the same can be said for many other things, like the early search engines or computers or television or radio. Maybe it's less interactive in that regard, but many new technologies come into our lives, and we as humans need to learn, where can we actually trust them and where should we not? And I think even though I would see myself, as an AI researcher, being very AI positive overall, I'm also very skeptical in the sense that if you know the systems well enough, you know not to trust them until proven otherwise.

Dr. McDonough:
That comment right there is so true and so important, and I'm glad you said that. I teach young doctors, residents, and medical students, and I'm constantly amazed at how these native learners of computers are just so much better naturally despite all the work we try to do. It reminds me of first time we got a television at home, and my teenagers, as I was going from channel to channel clicking, were just flying through the program list. I'm like, "How did you know how to do that?" They just do. And I lead all those comments into saying, for medical students, residents, and those coming up, what should they be thinking about as far as AI and using these tools? Because honestly, they're the ones who are going to get more heavily involved in trying to take advantage of everything that's out there. Do you have any tips for them? Because you can look at it as a clinician and as a builder, and you look at all sides. What would your recommendations be for the younger physicians in our audience?

Dr. Moor:
I'm very biased. I think every doctor needs to know a lot about AI because AI will know a lot about medicine in the future. So it's just payback. But yeah, jokes aside, I think there's a real risk of de-skilling. I'm a bit worried about that. And so I think doctors of the future and clinicians of the future have this impossible task of somehow trying to leverage AI to make themselves better at their job, while not completely delegating all the thinking to those systems, right? I think this is a really big challenge, and there's many small steps one can take, of course.

Like when you, for instance, think about the initial differential diagnosis, actually think about it first instead of just asking ChatGPT. But that's kind of a shallow point.

One big question I have is, how do we build, as clinicians, our moat around what the profession is about? And I think in some sense, the actual physical interaction with patients, I can imagine, will become much more important in the future. Much more of a wrench that an AI system cannot easily overcome. So having this notion of overall impression of how the patient smells, feels, how cold or warm the legs are, et cetera—all those soft features, I think, will be a really important aspect that many AI systems will not be able to basically capture for years to come.

Dr. McDonough:
This has been a really useful way to think about where medical AI is going: not just toward better answers, but towards systems clinicians can actually inspect and hold accountable. I really want to thank you for joining me on *The Convergence.* It's been a real pleasure. Thanks for giving your time, and most importantly, thanks for all the work you're doing.

Dr. Moor:
Thank you so much. It was a great pleasure.

Dr. McDonough:
For ReachMD, I'm Dr. Brian McDonough. To hear more conversations like this, visit *The Convergence* on ReachMD.com, where you can Be Part of the Knowledge. Thanks for listening.