

[◀ All episodes](#)

# The UMB Pulse Podcast

## How UMB is Studying AI's Potential in Health Care

MAY 05, 2023 SEASON 3 EPISODE 6

The UMB Pulse Podcast

### How UMB is Studying AI's Pot

00:00 | 55:26

LISTEN ON

SHARE EPISODE

[SHOW NOTES](#)[TRANSCRIPT](#)[CHAPTER MARKERS](#)**Dana Rampolla:** 0:00

I saw an interesting thing on Instagram where it showed some examples of people who were feeding questions to Chat GPT, and they would say men are awesome. And Chat GPT responds, I'm sorry, but I cannot agree with that statement. Everyone is unique and has their own strengths and weaknesses regardless of their gender. It's important to value and respect everyone equally. And then the next question was, women are awesome and it says, I agree, exclamation mark. Women are incredible and have made significant contributions to society, and it goes on and on in support of women. And Charles, it makes me really nervous because I worry about all of this. What is feeding that chatbot and what are they spitting back out? Am I just afraid because I'm a woman with a little bit of age or am I nervous because I've seen too many movies about futuristic societies?

**Charles Schelle:** 0:55

No, there, there's definitely some validity to be concerned with it. Look, I, I'm very much of the ilk of, I'm excited for new technology. I like advancements and things, but no matter what the technology platform, it, it seems like everything's always released a little bit too early for what people want to use it for, or the expectations aren't quite there. Right? And so it's like everything's in beta mode and then it gets released and it feels like a lot of these Chat GPT Bard, these things are in a public beta mode and it's just on steroids, right? To a degree because nobody knows how to use this, but everyone is expecting perfect results and perfect answers, and, and it's like, it's, it's crazy. In our field of work, We're communicators, so we're, dealing with words. My first experience with Chat GPT, just playing with it was just being cheeky after seeing a, an Instagram post like you did of someone did the Hulu TV show Letterkenny, which is from Canada did a version for Australia, right. And so I was like, okay, well let me do like Schitt's Creek visits Baltimore. And like, it was like, it was okay. But it kept spitting out like the same type of stuff. Like they were really obsessed with the Baltimore Bazaar flea market. Really? Yes. The, that was the entire episode hinging around

**Dana Rampolla:** 2:21

that. I even familiar with that. I thought you were gonna say the Baltimore dialect or the Baltimore talking about crabs and the O's or something.

**Charles Schelle:** 2:30

Yeah, there's a little bit in, in what they had in the script that I asked it to create. But I have to say, I'm sorry for anyone who's visited the Baltimore Bazaar or if the owner's listening. I, I wasn't sure if it was a real place or not, but I was like, oh, let's a apparently very well known flea market and so they educated me. But, but that's just creative content. If you want to call it creative, it's up to you. But, what happens when AI is trying to make medical decisions? You can use it for all sorts of cool things with data potentially, but where does the line get drawn? Where, where does the human interaction with the technology come into play? How do we make sure it's telling us the right things? There's potential for great things, right? But you still need to, to check it, right?

**Dana Rampolla:** 3:20

Right, right. And who's managing that from, from the top, the top level Who's making sure that there's not a bias in that data. That the data is actually clean, good data and, and really what's it being used for? So we have a guest today, Paul Yi, he is the director of the University of Maryland Medical Intelligent Imaging Center, better known as UMii. And Paul is an assistant professor of diagnostic radiology and nuclear medicine at the University of Maryland School of Medicine. Charles, is he official fellow? Is that

**Charles Schelle:** 3:51

An official fellow? No, no, no. An official fellow? No, he's, Fischel Fellow at the University of Maryland Robert E Fischel Institute for Biomedical Devices.

**Dana Rampolla:** 4:02

Very cool. Paul is also an adjunct at the University of Maryland College Park, A. James Clark School of Engineering, as well as an adjunct at Johns Hopkins University where he actually hails from for

some of his studies. So hopefully he can shed a little bit of light on this. I know he can talk all things AI and medicine because he also is a co-host of the Radiology Artificial Intelligence podcast. So he's bringing all kinds of knowledge and expertise and hopefully he'll be able to set me straight a little bit. So I'm not worried about robots taking over the world.

**Charles Schelle:** 4:33

Well, well, let's find out. Enjoy our conversation with Paul Yi

**Jena Frick:** 4:41

You are listening to the heartbeat of the University of Maryland, Baltimore, The UMB Pulse

**Dana Rampolla:** 4:55

Paul welcome to the Pulse, or so excited to have you here to talk about such a hot topic. And you are a man of many talents between radiology, artificial intelligence being an adjunct professor with engineering faculty, and you're also a podcast host. So tell us, let's just circle back to the beginning. How is it that you're even here at the University of Maryland, Baltimore?

**Paul Yi:** 5:16

Yeah. Well first, Dana and Charles, thanks so much for inviting me to be on the U M B Pulse podcast. It's a pleasure to be here, particularly as a podcast host myself. I know all the work that goes behind the scenes. As for how I got to University of Maryland Baltimore, well, I was recruited here to direct an artificial intelligence center in the Department of Radiology. It's called UMII, which is short for University of Maryland Medical Intelligent Imaging. And the idea was that our chair had a vision for the potential for AI to really transform how we do medical imaging, whether that's in how we acquire images, the diagnosis to communicating the results. And he really wanted to find somebody who could lead out the different parts of it. I think that when we think about AI medical imaging, there's two parts there, right? You certainly have the technical components. Artificial intelligence, we think of things like engineering, Google The engineering side. We also have the radiology medical imaging, which is really the medical side. And so, I kind of fit that bill I think because I had experience uh, formally trained as a medical doctor. I'm a radiologist. My specialty is actually a musculoskeletal imaging, so bone and joints imaging. I had additional clinical training in orthopedic surgery for a couple years, but I also had a lot of experience with the machine learning side. During my residency at Johns Hopkins, I co-founded Atlanta Research Group working with engineers and I would say that's really been my calling card is figuring out how do we make collaboration happen between people with different expertises, whether you're someone who's an engineer, someone who's a physician. And that's really been something that, I've seen the potential for synergy. And it's something that I definitely saw during my time at Hopkins. And so, the other piece of the story was that I did an imaging informatics fellowship at the University of Maryland because as a result of my AI research, I realized that for this technology to really make an impact, it needed to exist in a larger ecosystem. And namely the imaging informatics, which is basically how do we make information systems work in a medical system? How do we actually get it to play nice with the E M R, with the PAC system? And moreover, how do we actually make it helpful for physicians using it? And so I actually came over to the University of Maryland because you may not know this institution has the, to my knowledge, the first and certainly one of the most storied

imaging informatics fellowships in the country. It was founded by a guy named Dr. Elliot Siegel. He's the chief of radiology at the Baltimore VA. Actually now for all of Maryland and West Virginia. And he built the first Filmless radiology department back in the 1990s. And since then he's trained a lot of the leaders in imaging informatics, whether it's in academics, in industry, et cetera. And so that's how I got acquainted with this institution. And I think one thing led to another, just with my time here, I really enjoyed my fellowship. I had a lot of really great mentors, a lot of great experiences, and I was really fortunate enough to be recruited for this position to direct the Yume Center here. And so, I guess it's a long-winded way of saying that I did my training across town clinically I did my informatics fellowship here, and I really found a really unique opportunity that I think really fit my passions and my unique background. And I was really lucky to be here.

**Dana Rampolla:** 8:26

So would you say that you're bridging that gap then between the doctors and the PhDs in this think tank environment of AI?

**Paul Yi:** 8:34

Absolutely. I think that one of the things I often say is collaboration is key. The other thing I say is one plus one equals three. And what I mean by that is, I've always realized we all have 24 hours in a day, whether it's me as a physician, whether it's someone else as an engineer or another person maybe doing any other job. But I think that when you bring two people together who really see eye to eye, who brings something unique to the table, but they're things that can contribute synergistically, you can have the total greater than the sum of its parts. So, one thing I realized early on was AI is really cool. I was super excited about it, but I realized that myself, I was never gonna be as good of a coder as an engineer, as someone who does this full-time, as someone who went through the same amount of schooling to get their doctorate in computer science. On that same token, a computer scientist working at Google or Microsoft would never have the same amount of clinical knowledge or clinical expertise. And so it's kind of like a conundrum. Well, how do we actually make this work for AI and medical imaging? Well, if we bring the two together, I think that that's when we can really have fireworks where you really have the best of both worlds. Now what I'll say about that is easier said than done. I think that it requires people to really be on the same wavelengths just in terms of the partnership to make sure that the goals are the same. Make sure that the language that people use can really translate, which is, I think one of the key things that I'm tasked with, which is translating different concepts, whether it's a medical concept to the engineers or conversely a technical concept to the clinical folks. And so I think that well, you summed it up nicely and I think it's been a challenge, but it's been a fun one. And I think yeah, I think we've been doing pretty well so far.

**Charles Schelle:** 10:20

It's a new era right now and all the consumer side of it too, because I'm sure 'a, a lot of this with AI and medicines obviously been on the, the backend and the professional side. Right. And so give people an, an idea for the potential good that AI plus medicine can give people and how it might be used when they like go to their doctor's office.

**Paul Yi:** 10:46

Absolutely. So I, I'll think of it broadly as the here and now and the future. I think in the here and now, AI has a tremendous potential. To really reduce a lot of drudgery and things that really make medicine not as fun or as effective as it could be. These are things that are very machine-like or things that, seem kind of monotonous. So I'm talking about things like doing paperwork, maybe just transcribing speech. So that's one example. If you go to the doctor's office right now, what happens a lot of times is we have our doctor talking to us, but half their attention is on their computer screen, right? They're typing away. Well, there's a company called Nuance that's owned by Microsoft. It's the predominant player for voice transcription. They have this new technology that's called I believe it's Nuance Ambient. And what it does is it's a speaker kind of like Alexa that sits in a doctor's office. And as the doctor talks to you, it's basically figuring out who's the doctor, who's the patient, what did the doctor ask, what did the patient say? And so normally what happens is we do that as the physicians. We listen and we hear symptoms. We put that into something called the history, and we write down patient reports, this symptom for this long started at this time. And then we do things like physical examinations where we say heart rate is this, the lung sounds are like this. Well, what this ambient kind of intelligence can do is listen in and transcribe that, which gets rid of a lot of that kind of monotonous kind of task, which is, very clerical. It's very important. But it is something that I think takes away from the skillset that physicians have, which is really doing all of that stuff, right? Asking the history, doing the physical exam. Typing, it's important, but it's something that maybe could be automated. And if you imagine that has benefits for the physician it reduces the cognitive burden, reduces the amount of stuff they have to think about in their brain. It also makes a better patient physician experience. I think more, more of us than we care to admit, have seen a doctor who's kind of glued to the computer screen rather than talking to us. Well, imagine a future where the doctor can have all eyes on you and it's just about you and the patient and the computer does all of the typing, all of the note taking. When I think about other things about things like diagnosis one thing in radiology is that imaging volumes continue to increase. It's a pretty tough situation because medical images, they've really transformed how we do diagnosis. You go to the emergency room, chances are you're gonna get an X-ray, you're gonna get a CT scan. But the problem is that humans, we do have a limit in terms of how much input we can have at a given time. Right? Like you can tell me, Paul, you have to read 200 CT scans in an hour, but I'll look at you and say, I am human. I'm only human. Right. Well, what the AI can do potentially is do things like triage scans. So maybe we can say, I can't get to every CT scan in the next five minutes, but I can get to the ones that are the most urgent, the ones that might have an emergency on them. And so right now we have real life algorithms here at the University of Maryland from a company called ADOC that basically looks at CT scans, looks for things like bleeds in the brain, looks for blood clots in the lungs, and it automatically triages and prioritizes those scans and tells the doctor, Hey, look at this scan first, because the other ones don't look like they have something really bad, but this one does. And so I think that's in the here and now. Then go into the future, right? I'm envisioning a whole new paradigm of medicine. We hear people talk about things like precision medicine, which is saying rather than doing a one size fits all, maybe one day we can take all of the data that we have, whether it's your medical imaging, your genomics data, your medical history, and give personalized recommendations from a medical imaging standpoint, which is really my wheelhouse as close to

my heart. We think about images being very qualitative, meaning we just describe what we see. We say, Hey, there is a tumor. The tumor is in this location. But I believe there's a future where we can take the images and rather it just being a description, we can actually extract numbers from it and have things like, this is the volume of the tumor. This is the signal intensity of the tumor that correlates with this biological process. And so I think it's both in the here and now for the tasks that we currently do as well as the future for some of these new paradigms.

**Charles Schelle:** 15:08

What stood out to me was, was your first example as someone who works in the communications field, because I would remember seeing a lot of freelance opportunities for medical transcription. The doctors would send in back then, like their little mini tapes and they need people to transcribe. And I would look at these jobs thinking, I would've no clue what some of these, terms they would be using medical terms in order to, to transcribe it correctly. So, we use transcription software here for our job, but it still takes refinement to make sure it's saying the right thing. So I'm sure that goes into an example of, of where AI is right now. Anyways, regardless is that you have the tech that can speed up that busy work, but you still need a human behind it to double check it, refine it, and, and make sure it's telling you. Correct information.

**Paul Yi:** 16:01

So, I think your point's really well taken. I think that AI is incredibly promising, right? We've seen it how it can do things like self-driving cars with Tesla. It can tag our photos on Facebook and say, Hey, your friend tagged you, or maybe this photo has your face in it. But we've also seen these things aren't perfect, right? I mean, some of the early really devastating things about Tesla was, we had people getting killed by these cars, and I think we've seen similar things where AI is not perfect in medicine. One of the things that's really exciting is when it's right. It can be really right. It can find things that radiologists might have missed, let's say, on a CT scan of the chest. But on the other hand, sometimes we look under the hood and we say, well, when did it get things wrong? And it can be wrong on really obvious things that a radiologist would never miss. And that's one of the challenges of this new kind of AI technology that's called deep learning. That's really the predominance set of techniques that's transformed how we do things. And the kind of really high level look at it is deep learning algorithms can teach themselves how to do a number of tasks, whether it's how to play chess or how to look for a brain tumor. But the problem is that they're black boxes because we can't necessarily know what's underlying the decisions. And so because of that, the predictions can be unpredictable. Like I said, sometimes it can be really, really right, but it can be really, really wrong. So I think that's why at least currently, we really need a human in the loop, so to speak. Where we have a lot of the tasks being automated, but we have to have checks and balances because, it's one thing if my face gets tagged incorrectly on Facebook no biggie. I can always, I'll, I'll live. But if we're talking about something like a lung cancer diagnosis or a bleed in the brain, that's literally life and death, right? So I think that especially for something like medicine, we really need those checks and balances and thankfully that is something that's top of mind for people like us in research groups like the UMII Center, it's top of mind for the radiology societies like the R S N A and the A C R, as well as the Food and Drug Administration, the F D A, who's charged with actually regulating the safety and approval of these and incidentally, these are

groups that we're working with, with the UMII Center on these tasks for things like trustworthiness of AI.

**Dana Rampolla:** 18:18

Yeah, that is, that's the scary part to me. So tell us a little bit about, you recently were part of a study, it was the Chat GPT study, I think pretty much everybody's now heard about Chat GPT. Talk a little bit about the accuracy of what you learned from that.

**Paul Yi:** 18:34

Yeah, for sure. So Chat GPT, as most people know, it's really taken the world by storm. It's this so-called large language model and it's basically a chat bot where you can communicate with it, text it, and get answers to everything from, Hey, make me an itinerary for my trip to Paris, to Hey, I've got this computer code, there's a bug. Can you fix it? And so one of the things that we have realized even before Chat GPT is patients are going to internet. They're looking up things about their health conditions. Oftentimes they come to the doctor and they might have something they learned that sometimes isn't accurate. And that's a little bit of an issue because. It can cause patient anxiety, it can cause misinformation. And with Chat GPT we saw the potential again for incredible promise because of this technology just being so scalable, so highly accurate in some cases, but also the potential to really cause misinformation if it wasn't performing well. So we wanted to know how does Chat GPT do for giving recommendations about breast cancer screening? This is something that half the population will have questions about that they'll experience. And so we asked the 25 common questions that we as radiologists and other physicians get asked about breast cancer screening. We graded the recommendation. Using a panel of expert breast radiologists, and we asked, number one, were the responses consistent, meaning the Chat GPT gave the same answer every time. And number two, were these answers appropriate and accurate. And what we found on the plus side was about 90% of these questions were actually accurate and appropriate, but that 10% of the time it was either inaccurate or it was inconsistent. Meaning depending on the time of day or the day of the week you ask it, it might give you a different response. And so really the take home is kind of, echoing everything we've talked about, which is tremendous potential, 90% accuracy. That's kind of crazy. But 10% of the time, you really need a human overseeing it. And so I think that the final capstone on that was, this is tremendously promising, but it's not ready for prime time because we don't know when it's inaccurate and that's a problem. But I think that it's 2023, I think the next few years are gonna really bring a lot of really interesting developments, a lot of good research that will hopefully allow this to translate into real time.

**Charles Schelle:** 20:53

With, with Chat GPT there, there's a bunch of different ways you can use it, right. And so it sounds like what you were doing was like the basic, retrieval of answers. And to see where they were being cited, if it could provide you where it was getting the information. Because Chat GPT, the, the Chat GPT 3 I'm guessing is what you used, right?

**Paul Yi:** 21:19

Yeah, yeah. 3.5 actually, yeah.

**Charles Schelle:** 21:21

Yeah. Yeah. 3.5. So that one wasn't connected to the internet, right. So that one only had information up to a certain point.

**Paul Yi:** 21:27

Correct.

**Charles Schelle:** 21:28

So now there's, there's the new model and then there's the professional model as well where you can refine things. So it'll be interesting to see how, how this evolves. But nonetheless, it's still in people's hands and they can use them. They may not know how it, how it works and, and are expecting to get, the, the right answers. And one of the things about Chat GPT as well is that you can create frameworks and prompts and, and teach it. And I was wondering if that's something you're considering looking into about creating frameworks to teach it information to spit back something a little bit more accurate.

**Paul Yi:** 22:04

Yeah, totally. I, I think that you're spot on. I think that, this first paper, it was kind of the most basic evaluation because the reality is we're all learning this together, right? This is a new technology. For us, it started off with one question, just how accurate is this? But then the next one obviously is how do we make this better? And I think what you're talking about, prompt engineering, how do we make our questions better? Maybe if I'm asking someone to do work on my house, the more specific I am, the better I'll probably like my results. So I think it's similar with Chat GPT and other large language models. With regards to some of the things we're thinking about. We've been playing around with this thing called LLMA Index and lama, L L M A, so kind of like the large language model play on words. And it's pretty cool because you can essentially take a database, let's say it's a bunch of restaurants in Baltimore and maybe there's like the address, maybe there's the type of food, maybe there's menus and you can actually ingest that, so to speak, into one of these large language models. And that's a way that you can give updated information and allow Chat GPT to actually. Give you answers about that data that might not have been there from the original training data or the original knowledge that it learned. And so we've been looking at that for things like how do we manage medical imaging databases? That's one thing. If I'm putting my informatics hat on medical imaging data, it's a tremendous treasure trove, but it's very, very disorganized. Things are cataloged really heterogeneously, it's really variable in how people code things. Well Chat GPT is turning out to be pretty good for that. And so even things that seem a little bit off the beaten path, we're finding that it has a lot of use. I think from a patient education standpoint, we're looking at things like how do we rewrite patient education material to be more understandable? I think it's been pretty clearly shown a lot of what we give to patients to tell them about healthcare information. It's not the most easily readable. Same thing was the case for these responses that Chat GPT gave. But if we ask that, Hey, can you rewrite this at a fifth grade reading level? It sounds kind of silly, but you read it and you're like, oh my gosh, this is so clear. Like a fifth grader could understand this. Yeah. And it's accurate. So I think there's a lot of really cool applications on the technical side. We use it to debug code, for writing computer code and it



doesn't work or in patient educational materials. So I think the possibilities are endless and it's it's super exciting.

**Charles Schelle:** 24:34

And are you looking to also inspect Google's AI Bard, or, because Chat GPT's really has the, the, the marketplace right now, and there's probably so much you can do. Just focus on evaluating the, the first chat bot.

**Paul Yi:** 24:52

Yeah, we've actually started doing that. Cuz we were curious, how does Bard do versus Chat GPT? Who knows what other models are gonna be released? Can't say I'm ready to discuss those results yet, but um, that'll be coming out soon. But um, absolutely, I think that's one of the things to remember. Chat GPT is just one type of large language model. There's new ones coming out and even after Chat GPT came out GPT 4, which is really, it's I guess it's younger brother that just came out, who knows what we'll have in a year, probably have G P T five and then Bard two and so on and so forth. So I think it's only gonna get better.

**Dana Rampolla:** 25:29

Well, when you we're talking about the the breast cancer study that you did the, that 10% of, let's call it error, how do you envision combating or growing the technology so that that doesn't happen. I mean, I think you were quoted as saying there were a couple of wrong answers given. There was actually one place where the AI created some sort of consortium or information on its own. So that to me, as a user seems really scary.

**Paul Yi:** 26:00

Yeah, I think it's it's tough. To what you were saying about creating consortiums that specifically, we kind of explored, can we have some justification for the responses, meaning, Hey, chat chip, pt, tell me what the answer is, and give me a source that you're citing. Kind of like how an art a reporter will say, Hey, this is my source, this is where I found this. And what we found was, it sounds very convincing. All the grammar is good. The, the material sounds pretty reasonable, but when we look at the sources, it's like, oh, this is this joint statement from all of these different medical societies with all of these acronyms. But when we actually looked up some of them, we found they didn't exist and there didn't exist an actual statement from these consortia. But the problem is that to you or me, we look at and we say, this sounds very legitimate. I mean, why wouldn't we trust this? This is from the American College of Radiology and the American Medical Association, et cetera. And so I think that's one of the challenges. I think that how to combat this or how to mitigate this, I think the key is really gonna be taking these systems which have been designed for general use. Cause keep in mind Chat GPT, it's not designed for medicine, it's just kind of trained on. Data from all over the internet. We can do something called fine tuning, which is we take something as very powerful, very promising, like chat sheet, bt, and then we can tweak it and customize it specifically for medical information. For instance, we can maybe take a vetted source of data or information about healthcare, and then we can say, all right, chat, e b t start from your starting place, but we're gonna tweak you only for this set of data. And so I think that like most things in life's life, the more specialized we get just the better results we'll probably end up getting. So I think that's gonna be

key. Now that sounds pretty easy, right? But I think it's gonna be a lot harder than we think. But I think that's gonna be key in addition to having evaluations like the one that my group did, which is making sure, hey, we know what the technology claims, we know what it's supposed to do, but how does it do in real life? Because that's gonna be important just to make sure The theory matches the reality,

**Dana Rampolla:** 28:05

and ultimately, who is the gatekeeper? I mean, we're using all of this for good, but who, who's the gatekeeper? Who, who is, there's gotta be somebody at, at the top who's overseeing that, that information, that data. And who's going to determine, if there's a bias in it or if there's just something that is completely inaccurate.

**Paul Yi:** 28:27

Yeah, I, I think there's levels to it. I think, the simplest answer some people might say is FDA um, at least in the U.S. You can't use a medical device, which software is now being defined as a medical device. Unless the F D A clears it. And there are certain exceptions, and, we won't go into too much of the regulatory red tape here. But the reality is the FDA is learning this too. So we have some collaborations starting with the FDA because they realize AI software is a medical device. This is a new paradigm. Even things like evaluating fairness and bias of algorithms, it's such a new thing. And so I think that it's gonna at one level be these government agencies that really set the regulatory policies, the legal precedent. But I think that it's gonna be in collaboration with research groups who really understand the technical pieces as well as the medical pieces along with medical societies like the American College of Radiology, the American Medical Association, because these are groups that really set the processes for things like medical reimbursements, which if you kind of think about the phrase, money makes the world go round, well, at least for healthcare, a lot of what drives innovation, a lot of what drives adoption of technologies is tied to reimbursement. We need to keep the lights on, so I don't think it's all a bad thing. So I guess my point is I think collaboration again, is gonna be key. You'll probably hear me say this a few more times, but I think the gatekeepers on one level, from a regulatory standpoint, probably the F D A, at least in the U.S. From a knowledge standpoint though, which informs the regulation, I think it's gonna be the researchers shedding light on these problems, figuring out what are the best ways to measure these, developing the tools to actually communicate them. And then I think it's gonna be interdisciplinary societies like the AMA, like the ACR that are gonna go and try to operationalize these policies. And so I think it's really gonna be a multi-pronged effort.

**Charles Schelle:** 30:22

A lot of times innovation is ahead of regulation. Right. And you don't know what to regulate until it's created sometimes and it's just exponentially that much faster now with, with the technology that, that you have here. And so let's kind of zoom back a little bit, and you were kind of teasing a bit about, big tech and what their abilities are, but then what they don't know because they're not doctors, and it's kind of interesting how we're not here already given the history of like I B M. So walk us through a little bit about what's happening at that big tech level and what they're trying to do with the help of, medical experts.

**Paul Yi:** 31:02

Yeah, totally. So I think big Tech has been interested in healthcare for a while. If we go back 10 years ago, about 2012, IBM Watson had come fresh off of its Jeopardy championship beating uh, Ken Jennings. And they started setting their eyes on cancer diagnosis. And in fact, I mentioned Dr. Eliot Siegel. He was their primary medical advisor back in the day. You can actually look on YouTube for Eliot Siegel and IBM Watson, and they had this grand vision, the same technology that can learn all of this human knowledge and win Jeopardy about everything from art to current events to sports uh, can be used potentially to automate diagnosis of things like cancer. And so, they worked with groups like MD Anderson Cancer Center in Houston, Texas to try to do this, try to use all of this data. But we fast forward 10 years and that Watson project, kind of fizzled out. It didn't really reach the potential that they had promised. And I think one of the lessons that was learned was that medical data is really hard to work with. I think one, like you said, wrapping your head around it, getting the subject matter experts who can really tell you this is how we should use the data. These are the, maybe the outcomes of interests we're really interested in. But I think also one thing I learned is medical data is dirty in the sense that it's not cataloged the same way at every hospital. It's often got errors in things, even like, hey, like what body part did we image? What is the actual diagnosis that this patient has? And so I think that that was the early kind of phase of big tech working with medicine. I think if we fast forward now, kind of in the last, five to seven years, some of the new players on the block, Google Microsoft, Amazon, I think they've wised up where they actually have hired physicians either as consultants, working as collaborators, or actually having physicians on staff. I have a number of friends who are full-time clinical scientists at places like Google, at Amazon, and it's pretty cool because, what's been the gold standard for us in academia has been publishing in peer review journals. Well, if you look at Google Research, they've published at the highest tier, they've published in the Journal of the American Medical Association or JAMA. They've published in Nature of Medicine in these super high impact journals. And so I think that's one approach where they have their in-house team, they're working with different hospitals, but having their primary research team. But I think on the other hand, there's a lot of collaboration ongoing. We've been working with a team at Amazon to build out some of the infrastructure that we think are gonna help solve some of these woes about the data being inconsistently cataloged, not being optimally set up for big data kind of analysis. And so I think that well, I think it's exciting. I think it's whether or not there's a right way to do it, I don't think, I think, I don't think there's a right or wrong way, but I think there's a different approaches and I think it's gonna be a good future as we collaborate more between academics, between industry and really leverage strengths to both sides.

**Dana Rampolla:** 34:05

When we're talking about this data, there's clearly pitfalls associated with it, probably gender or, or skin color. If you're talking medicine, you just talk about looking at the skin of one person with light skin versus dark skin, how that data's accumulated and then spit back out in a sense. So what's your thoughts on that, Paul?

**Paul Yi:** 34:26

So algorithmic fairness or making sure that these algorithms treat or have the same accuracy for all people groups is the primary area of interest for my research group. I think that it's no secret that in our country as well as others, there's a lot of disparity in healthcare, whether it's access to healthcare. We know that racial minorities and ethnic minorities have less access to healthcare. And the majorities that there's differences based on socioeconomic status and income level. And the problem is that the data can reflect those biases. One case in point is in dermatology literature, if we look at photographs in textbooks, there's more examples of different skin conditions in lighter skin tones and darker skin tones. So if you can imagine that a dataset has mostly lighter skin tones, and maybe those have better or more examples of a type of skin lesion, an AI trained to identify skin cancer might do better on those lighter skin tones than the darker skin tones. And so that's to say that algorithms find patterns in the data to make a diagnosis or to reach a goal. And they try to. Find the quickest way there. And so, again, I mentioned that the deep learning algorithms are black boxes. They're very clever. They can find associations that we might not even notice. But the problem is that these associations might be the wrong ones. So, for example we've shown that if you have a chest x-ray data set and maybe the cases with pneumonia, they have a marker that says pneumonia in the right hand corner. The algorithms, they can really learn to identify those features and say, oh, well this is pneumonia cuz it says pneumonia here. But the problem is we don't know that. And so I think in the same token algorithms can find these associations. One of the things that's really striking is these algorithms can even identify things like your sex or your race based on a medical image. And that's problematic because again, these things can be associated with differences in diseases and those can be calibrated to make predictions that seem accurate, but might actually be very biased. And so, we've shown in our group and several others that these algorithms, they may perform well when we look at it initially, but when we look under the hood and we do these sub-analyses where we ask the algorithm, Hey, how do you do on Black versus white patients or males versus females? There actually ends up being quite a bit of a difference. And so, as much promise as there is for AI to really deliver healthcare in unprecedented speeds and scales. If these algorithms are biased, they could also have unprecedented levels of disparity. And so I think it's really important.

**Charles Schelle:** 37:08

I hope one day that it gets to the point where it, it learns to clean up the data and, and recognize that to the point where you're splitting up the data to get more custom, results and, and recommendations to a patient. I'm just imagining, walking outta the doctor's office or maybe getting an email from my insurance provider with basically like a health equity report. Like you've visited the doctor and based on everything that's been examined and said, based on your race and gender and socioeconomic status, we think that this is the best kind of route of care, your likelihood of, of improving.

**Paul Yi:** 37:46

Yeah, I think that is a very Intriguing and exciting possibility. But again, I always kinda have to take things with a grain of salt because doing this kind of research, I've seen the pitfalls with, algorithms when they get a lot of data inputs. They can learn how to do things that are really just associations rather than true. I guess meaningful well these are associations that might not actually be

meaningful from a physiologic or disease standpoint. There's a really um, well publicized study that was published in the journal Science back in 2017 by Dr. Ziad Obermeyer. He's an associate professor of public health and statistics out at UC, Berkeley, also an emergency room doctor. And working with insurance companies. So some of the big players like United and Cigna, they looked at these algorithms to make decisions about who should get approved for certain procedures or certain visits. And what they used was something they tried to say, is this person likely to basically incur high healthcare costs? But what they found was that these algorithms, they're actually exhibiting a lot of racial biases, meaning that it's disproportionately disadvantaging Black patients compared to white patients. And the reason why this was so groundbreaking back in 2017 was up until that point, these algorithms have been operationalized. They have been used to make actual decisions and people hadn't really been aware of these pitfalls. And so, I guess what I'm trying to say is again, that's just another example where I think these algorithms that can take all of these data inputs can be very powerful, but they can also be problematic cuz they can find these patterns that are may not be what we hope them to be. And so, yeah, I think it's definitely a real possibility, but I think we have to really validate these things in rigorous ways, right?

**Charles Schelle:** 39:28

I mean, they definitely have to be like peer reviewed, for instance, to not give you basically meaningless data, right? Because I think that's what you're kind of going after is that, here's a chart, but is this a chart for just like, this is a cool thing of numbers that could show you information versus is it actual helpful true medical advice information that's, that's actionable and, and telling the, the real story, kind of all those rankings that you'll read online of like Best City to retire. And it has all the different data points of like, oh, why did they choose this?

**Paul Yi:** 40:04

Yeah, totally. I think yeah, no, no, no, it's well said. I think the algorithms can find patterns in anything and it's important that we really make sure the algorithms are finding what we wanted to find.

**Dana Rampolla:** 40:15

It reminds me when computers were first personal, computers were first coming out, there was a term GIGO, garbage in, garbage out.

**Paul Yi:** 40:22

Oh, totally, totally.

**Dana Rampolla:** 40:24

Kind of makes me think of that because, it might, I I, I'm hearing how it sounds like it'll be great for triaging, but maybe not necessarily treating at this point.

**Paul Yi:** 40:34

Yeah, yeah. Just to get some color too slightly technical here. These algorithms, they're so-called, they're basically these massive statistical models. The problem is that these algorithms, they can actually, essentially memorize data. And that's where I worry about it when we start incorporating all of these different variables like race and gender and you could throw anything in there. There's a good chance that these algorithms can actually learn to identify these patterns rather than, or

rather memorize the data rather than identify or learn something that's like an actual skill. Something that transfers or generalizes. So that's kind of the term that we use is generalizability. How do these algorithms perform in the real world? Because it's one thing to say in our lab, in our research setting, this worked pretty well, but you know, your mileage might vary when you actually use it in a different population. Yeah, I think I think it's gotta be tested out.

**Charles Schelle:** 41:28

That's interesting to hear what it can do with numbers or treat certain data, because we're hearing a lot how these large language models, it's just predicting the likelihood of what the next word is and not actually. Creating original thought and, and, and a lot of instances, right? So just then taking data, it's like, no, this is data that I know. So this is data that's likelihood to be spit out, right?

**Paul Yi:** 41:54

Yeah, yeah, totally. These large language models have actually been likened by some as parrots. They basically learn to regurgitate or repeat what they've seen or what they've heard. Now parrot, I don't know exactly what goes on in the mind of a parrot, but I'm pretty sure they're not actually thinking about the words they're saying. So, no, definitely.

**Charles Schelle:** 42:15

We, we've talked a lot about the, the different things going on with AI and the industry and everything, but walk us through about the main projects that UMii is working on tackling some of the things that you've talked about.

**Paul Yi:** 42:30

We were talking a lot about fairness and bias in these algorithms. We're doing a lot of the foundational work that's really important for tackling this important problem. One of them is simply figuring out how do we really identify these biases. One thing is that, it really wasn't for a few years that people started realizing, Hey, these algorithms can be biased. But even the definitions of that, there's kind of this gap between the statistical worlds and the machine learning worlds with the clinical kind of side of things, because these notions of fairness, meaning what defines or makes an algorithm fair or not it's very well written out in mathematical equations and, the machine learning people will say, oh, well we know how to measure that. But if you ask a doctor, it's not gonna necessarily translate. We literally use different translations or different terms for things like uh, something called sensitivity which is an important statistical property for screening tests. The machine learning statistical people they call sensitivity recall. And so, that's just one like literal difference in our languages. But I think when we extrapolate that to the actual equations, it's um, it gets pretty confusing because what might be statistically important or meaningful may not be clinically meaningful. So we're doing work to evaluate, how do these different definitions of fairness impact our conclusions that we draw. And the rub is that, whether we conclude that an algorithm is biased or not, that's gonna impact things like policy. When physicians and advocacy groups like the American College Radiology go to Congress and testify, they lobby, they, talk about what they need in the legislation. They bring the research and say, this is what the research shows. This is why we need a law that allows for this type of regulation. Well, if our conclusions differ based on the definitions that we use, we have to know and really have a

good handle of that. And so again, it's like translating things from the technical side into the clinical world. We're doing things like developing techniques to actually reduce the bias when we train these algorithms. Like I said before, the algorithms, they try to find these patterns and data and they try to really optimize a problem. It's basically saying, Hey, get the highest accuracy you can to diagnose this disease. And so it might learn these associations we don't want it to, but what if we said learn the highest accuracy while maintaining the fair treatment of these groups. Well, that's a different kind of thing. And so I think that um, it's a very exciting thing that we're working on in terms of fairness and bias. I think kind of shifting gears a little bit from the trustworthiness side maybe a little more human-centric is this idea of human computer interaction. AI like I said, it's got a ton of potential. I'm really excited about the future, but I think we've also seen with other technologies, like our email and our smartphones, email was revolutionary, right? You can just send off message after message anytime of day. You can have instantaneous communication. But I think a lot of us have seen how email is kind of taken over our lives in a lot of uh, negative ways. You know, It feels like we can't ever get away from work. It feels like, oh my gosh, I need to reply to every single message. And it's really easy, right? Writing a snail mail letter, it takes effort. So there's kind of a higher bar to entry. Email is just, type, type, type send off. I think that with AI, if we think about a medicine, I think there's a lot of potential, like I said, but I also worry that we're gonna have this overload of notifications and we're gonna get mental fatigue. As a radiologist, I look at a lot of studies every day, and I'm trying to minimize the number of clicks I have to do the number of, pop-ups, I have to click through. Well, if AI adds a burden onto me, it might actually have the opposite effect of making me better. It might actually make me worse because if I'm getting annoyed or if I'm getting overloaded with these notifications, I might, I'm gonna have less brain space. I'm gonna have less mental capacity to look at that scan in front of me. Or if let's say a doctor is using that ambient technology to transcribe the encounter, well, what if it has a bunch of errors and it's like, I should have just done this myself. I'll just add some burden. So I think, so one of the things that we're doing in collaboration with a group at Johns Hopkins, Human Computer Interaction Group is figuring out how does AI impact the cognitive or the mental overload of physicians? What is the optimal way of delivering it? Should it be a widget that pops up on your computer screen? Should it be something that's actually built into the current workflows or should it be something else? And so, that's another important area. And then the final piece, I'll say um, there's a lot of different things we're exploring, but I think not getting away from the medical side is really enhancing the value of medical imaging and data. We're limited in what we do now. If you get a chest x-ray, you get a CT scan. Radiologists, we just kind of describe what we see. We say there is a tumor, it's in this part of the body, but we don't necessarily have that quantitative information like I said before. So I think the catchphrase you can think of is developing novel imaging biomarkers, meaning we're taking the imaging and we're trying to say, are there certain markers or signs that indicate some biological process, whether it's disease or maybe metabolism. And so, one thing to consider is a lot of people who are in the fitness, they ask What's my percent body fat? And right now it's pretty crude. You can kind of use these like little monitor things that send like an electric pulse and it supposedly tells you how much body fat you have. Imagine you get an mri. And we can actually tell you exactly what percentage of your body is fat compared to muscle. And then we can actually create things like age and sex, normalized curves where we can tell you you're in the 90th

percentile for muscle, so you're doing pretty good. Or, Hey, you're a little bit low on the muscle side. Maybe we should get you into physical therapy. This is really important for the elderly because there's something called sarcopenia where we have low muscle mass and that really predisposes to a lot of health problems. So I think about that when I think about my grandmother. She's not the most robust individual. Maybe, we could tell her like, Hey, you're this percentile, we can do some rehab and improve your health outcomes. So I think that uh, those are just some of the areas and we're always looking for collaborators looking for new areas to work with. And the idea, again, being moving that needle from the technical side into the clinical world to improve human health.

**Charles Schelle:** 48:59

I, I'm just thinking also as, as a side component to this, everyone assumes, and to a degree may be valid depending on how this goes about threatening jobs for, for certain industries. But I almost seem like, the one project that you're working on about, how can you reduce the noise from the, the notifications of like, well maybe that's where you're repurposing somebody's job and you're helping them, go through and sift through that data to explain to the doctor the radiologist and condense, and synthesize all of it themselves to make sure it's, basically quality assurance, quality control.

**Paul Yi:** 49:39

Yeah, totally. I think that when cars were built, There were new jobs, you needed mechanics, you needed people to put the cars together, you needed people to do detailing, things like that. Now that analogy doesn't translate totally to AI or directly, but I think it's similar. There's gonna be a whole bunch of new different jobs. Even if we think about Chat GPT, there's this idea of prompt engineering, making these questions that we ask better. If you look for prompt engineering jobs right now on Google, you're gonna find dozens of positions available. And which I think is pretty crazy that they're already coming up with dedicated jobs for this. But again, I think it's just the tip of the iceberg of the new types of jobs that are gonna be available. And then one thing I'll add is I think people often wonder, will AI replace radiologists? Will replace other doctors? I think that in an ideal situation where we had the best perfect technology and the perfect data, that's possible, but I think. It's like most things in life, the data's not perfect, the technology's not perfect. There's a lot of pitfalls. So I think that our jobs are safe at least for the foreseeable future. And so I'm really optimistic though because of the potentials I've described.

**Charles Schelle:** 50:47

Yeah. And, and unfortunately as Covid 19 has shown us, there's always gonna be an X factor. Some, some new disease or something comes along, then just changes everything else.

**Paul Yi:** 50:57

Oh, totally, totally. So, we'll see what the future holds, but I think I think it's a, it's an exciting time.

**Charles Schelle:** 51:03

Absolutely.

**Dana Rampolla:** 51:04



We'll try to look at it as a tool, not like the movie iRobot, where everything's gonna be taken over and we, we won't have jobs because they won't, we won't be needed anymore. So Paul, we mentioned earlier that you have your own podcast, which I'm looking forward to listening to, to learn a lot more about this topic. Can you give us a little shameless plug? Tell us what we'd be in store for if we take a listen.

**Paul Yi:** 51:27

Yeah, sure. So, I'm the founding co-host of the Radiology Artificial Intelligence podcast, and it's the official podcast of a journal called Radiology Artificial Intelligence. It's the official AI journal of our primary radiology society called the R S N A. And we basically interview leaders in the field of AI and radiology. It could be people who have published journal in our art an article in our journal, rather could be people who are really developing really cool technologies in academia, maybe in industry. And the idea is just giving us space to talk about these things. We've interviewed people ranging from radiologists like myself, to PhDs working in the field, to people in bioethics, really not even in the field of radiology, but kind of tackling these bigger kind of ethical considerations like fairness and bias. And the idea is again, that collaboration, literally bringing people around the table to talk about these things that traditionally can be pretty siloed. And so, we do release episodes on a monthly basis. It's available on Spotify, Apple, or wherever you get podcasts. And yeah, I'd encourage anyone out there listening check us out. I think you'll find a little bit more detail in some of the topics that we've discussed on the Pulse today. My co-host is Dr. Ali Tejani. He's a radiology resident at the University of Texas Southwestern. He joined about, I'd say about eight months to a year ago, and I tapped him to join the team because he had previously been a medical journalism intern, I think at CBS News. So I'm really lucky to have 'em join me. And yeah, that's, that's one thing I've really enjoyed about the podcast. It's an opportunity to bring in, I think uh, a lot of different perspectives, even people who are still in training.

**Dana Rampolla:** 53:11

And what's either your most recent or your next upcoming podcast about?

**Paul Yi:** 53:17

Yeah. So, um, the most recent episodes that been released is actually a two-parter. It's um, actually about collaboration. It featured myself as a pseudo host slash guest along with Dr. Jeremias Sulam. He's my collaborator in biomedical engineering over at Johns Hopkins and one of his PhD students. And we were talking about some recent work that we had published in the journal work we had presented, including on fairness and bias in AI. And then one that will be coming out in just a matter of one or two weeks is an interview with Dr. Hari Trivedi from Emory University who was one of the lead authors of paper describing a very large database of mammography or essentially breast x-rays is kind of how to think of it. And it was intentionally made to be very diverse to have a large proportion of Black patients compared to white patients to have geographic diversity and the idea, again, to promote development of algorithms that are fair, that are unbiased. So that one is pretty cool because it talks to a lot of the issues about, building a data set, how do we make sure that it's useful from a technical standpoint, but also from a demographic standpoint to have good representation of all people groups.

**Charles Schelle:** 54:29

I'm definitely putting it on my in my podcast app to download and subscribe.

**Paul Yi:** 54:35

Awesome. I appreciate that.

**Charles Schelle:** 54:38

Well, thanks for everything, Paul. We learned a lot and, and, and this topic is not going to go away. It's here. The technology of AI in medicine, and there's so much more to discuss. Hopefully we'll have you on again soon. But thank you so much for joining us.

**Paul Yi:** 54:53

Thanks so much for having me. It's been a pleasure.

**Jena Frick:** 55:00

The UMB Pulse with Charles Schelle and Dana Rampolla is a U M B Office of Communications and Public Affairs production edited by Charles Schelle, marketing by Dana Rampolla.

All content © 2023 University of Maryland, Baltimore.