This transcript has been edited for clarity.

John Whyte, MD: Welcome, everyone, to Medscape Masters. We're excited to have you here for tonight's discussion, "AI Is Here. Is Your Practice Ready?" We're joined by Dr Eric Topol, Dr Raj Manrai, and Dr Ted James. I'm Dr John Whyte, WebMD's chief medical officer, and I'll serve as tonight's moderator.

Q&A is an important part of the event. Some of you have sent questions in ahead of time. I'm going to try to incorporate them into some of the questions I'm going to ask upfront, but I also encourage you to enter any questions you have for the experts into the question box, which is to the right of your screen.

I'm excited to introduce our experts today:

Dr Eric Topol is Medscape's editor-in-chief, a world-famous cardiologist. He has covered AI in medicine in the GroundTruth's newsletters and in Medscape's *Medicine and the Machine*.

Dr Raj Manrai is the deputy editor at *NEJM AI* and is also an assistant professor of biomedical informatics at Harvard Medical School.

And Dr Ted James is the medical director and chief of breast surgical oncology at Beth Israel Deaconess Medical Center and associate professor at Harvard Medical School. He has written and spoken extensively about AI and its significance to physicians.

Doctors, thanks for joining me today.

Ted James, MD: Thank you.

Arjun (Raj) Manrai, PhD: Great to be here.

Whyte: Medscape just published a [report] based on a survey of how US physicians view the future of AI in a medical workplace. I just want to review a few of the key findings. The first is that more doctors when we look at the numbers, are enthusiastic about the future of AI in the medical office than are apprehensive or neutral, which seems encouraging; you can see the numbers here. But how do physicians expect their organizations to use AI in the future? Doctors most often think their offices would use AI for office administrative tasks, in patient and staff scheduling — for example, using ambient voice technology to create notes during a patient meeting. But around two thirds of physicians also said they expect their employers to use AI to predict prognosis and diagnosis. I think it's interesting to anticipate where AI will take hold in physicians' offices first. And let's talk about some of those administrative tasks. How important are they going to be? How long are we going to keep with them? Eric, you and I have talked a little bit about AI's potential value here. What are your thoughts about the role in administrative vs deliberative decision-making processes of diagnostic treatment?

Eric Topol, MD: John, good to be with you. The "administrative" makes it sound like it's not so important, but as you touched on, it's about being able to capture the notes, the synthetic notes from the conversation between the patients and physicians, adjusted by articulating what the physical exam shows, which otherwise might not happen in the

conversation. With that simple adjustment, not only can better notes be generated than what we see today in our health records, but also all the downstream tasks, whether it's a follow-up appointment, scheduling, lab tests, procedures or preauthorizations. And the list goes on, including nudging patients about things that were discussed. And that note is referred back to the audio so that if there's any confusion, the patient can check what exactly was spoken. This, I think, represents the future. It's not a little administrative thing. It's actually keyboard liberation — marching toward keyboard liberation — and getting rid of data-clerk function. We've already seen it taking hours of work from many physicians' practices. So that's a big thing. Now, of course, there are many other features that were in that list, but I just want to emphasize that administrative is more than billing stuff.

Whyte: Ted, I wanted to ask you how your colleagues are thinking about this, because sometimes you talk to physicians and they're a little bit reticent or concerned that everything they say is documented somewhere. We've been through earlier iterations, remember, where we talked about recordings. If a patient ever says, "Can I record the conversation?", sometimes there's a little bit of reticence about that. So, what are the current thoughts on that? We did have a question ahead of time in terms of how well these AI scribes or other systems are performing. Are they performing well? Because we've all seen in dictation, for those of us who are old enough, how sometimes there are some big errors.

James: Overall I think there's a lot of interest in AI. I'm very happy to see that, because I think it's going to be transformative. Obviously, there are things that we need to do to put in safeguards in terms of data security and transparency, in terms of avoiding any biases in the AI and errors. Obviously, that's an important thing. But the fundamental power that AI brings is going to be really transformational in terms of what we can do for tailoring care for our patients, and, as Eric was saying, alleviating the burdens for the physician so they can focus more time on patient care, which is really what we want to do. So, by and large, I'm happy that there's a lot of interest. I think there is a lot of trepidation, though, because as you said, John, other technologies have been introduced that didn't necessarily alleviate burdens but, if anything, created more tasks and steps for physicians. So there is some potential apprehension there, but we can learn from past mistakes that we made and really look at AI as something that's going to be a value-add to clinical practice. By and large, there are many physicians like myself who are very optimistic about its potential. You had mentioned the use cases and the digital scribes as one of the ways that AI very early on can make a big difference. We have some organizations — I believe Moffitt has published their results — and they find that these digital scribes are acceptable, accurate, and relatively easy to use. And these are the things that will alleviate a lot of burdens, decrease burnout, and, as Eric said, make things better for both physicians and patients.

AI Pending Orders: Ready for Prime Time?

Whyte: Someone also asked, "Can these AI tools pend an order or is that a step too far?"

James: Raj, Eric, you've got experience with that too. I think that's not quite ready for prime time, from what I've seen. Raj, do you have a different thought, perhaps?

Manrai: We are seeing early glimpses and efforts of tools that are designed to try to do that. But as Ted is saying, I don't think those are yet ready for prime time. There's the technology and there's the implementation of the technology. You see the technology itself, which is the actual digital scribe, a computer-based or phone-based agent that's

listening in on the conversation and allowing the patient and the physician to maybe actually make eye contact, which they haven't done. I'm hearing from some physicians that they haven't been able to make this much eye contact in a decade since they've been practicing. But now that they have their AI that's listening in and writing the note, and then they're checking it over, they're actually able to engage with the patient in a much more meaningful way. So I think that technology is impressive, but the integration into the EMR, with other notes, with a patient's full medical history, those challenges are still there for the major players.

Whyte: Raj, how will that work? Because we can't capture the entire conversation with that note, right? That's too much information. So are these tools getting better at figuring it out? Like you said, they're not quite ready for prime time. What's missing? Because they're going to get better.

Manrai: They are getting better. I think we're piloting them now. But as Ted referenced, there are studies that are ongoing. I am optimistic about their capabilities based on what I've seen, based on what we're seeing. They're already quite good. They are recording or transcribing the conversation, but they're also summarizing it. They're also putting it into a format that is accessible to the physician later on and understandable to the patient. I think that is getting better rapidly and it's going to keep getting better.

Whyte: So, none of you want them to pend an order. I want to ask about that because we do have our medical staff, administrative and nursing staff, medical MA staff; they'll often pend orders about mammograms, about what's due in terms of immunizations, or a diabetic foot exam. But you don't think that's ready for prime time? Anyone want to jump in there about being able to pend orders that would reduce administrative tasks.

James: I think we're really close. The capabilities continue to move at an exponential rate. They have these multispeaker transcriptions now and they can take large amounts of spoken data and process it quite well. So I think we're very close to the point where orders will be pended. But I don't think we should get to the point where the AI should replace physician checks and balances. The last step is always going to be the physician. But I agree with you: I think soon we will have an AI that goes far beyond just transcribing, maybe pending notes and orders that would have to be reviewed by the clinicians. There still is the issue of false information and word-error rates with AI that needs to be hashed out before they can be 100% accurate. So it is going to require physician oversight to make sure that the information is correct and the orders are accurate.

Whyte: Eric, you were going to say something about the pending orders.

Topol: I think it's ready. There are some health systems that have already universally adopted it after pilot studies, so this is going to spread pretty quickly. The hallucination or confabulation problem — it isn't like when you're doing a prompt generally to a large language model; it's just using the conversations that took place. It's doing a much better job than most doctors do of formulating their notes. And as you're getting at, John, the other point is trust, because you can refer back to the raw conversation and it's all linked. So whether it's the doctor — and I agree with Ted; it requires oversight — or the patient, it's all recorded there. And this eventually will be the norm. I think we have to adjust to it because it's going to save so many hours of data-clerk function, which is what every clinician hates.

Whyte: I wanted to bring up the fact that AI could help write responses to denial letters. There's actually information in the news that insurance companies are using AI to write the response that a previous AI wrote to the denial. Ted, what are your thoughts on that? Are we getting too far ahead in some areas?

James: That's an interesting thing, and I think we are going to see it on multiple sides and multiple levels. But the point of the matter is that many of these repetitive, burdensome administrative tasks that are taking physicians and clinicians away from patient-facing time can be automated. That's really where I think the promise is. As much as we can automate these tasks and make it so physicians don't have to be spending their time doing that, it's going to make things better. Now, I always say that the issue of physician burnout is not necessarily so much about our being busy, working hard, or working in stressful conditions. I think that's what we signed up to do. It's all of these hurdles in our way. I say that it's not the long journey but the pebble in the shoe that wears people down. So the more that we can automate things and remove these pebbles, the better it's going to be for clinicians. Whether or not the insurance companies use them as well… We want to alleviate the effort for physicians.

Is AI More Empathetic Than a Physician?

Whyte: One of the questions that came up, Raj, was, do we expect to see a AI system or AI assistant embedded in the EMR? When you talk about those repetitive tasks, one of those repetitive tasks that we often have is answering questions from patients, for which there is remarkable consistency over the years when you practice for a long time. Do you think we'll be seeing that embedded in the EMR as an AI assistant?

Manrai: I do think we will see that more and more. Many health systems are already piloting versions of that. There's compelling evidence from the scientific literature. There is almost a narrative, right? When GPT-4 first came out, it could do some amazing things, but it would never replace the empathetic communication that you can only get from a physician, from a human. There was a paper that was published shortly after that evaluated the notes generated in response to patient queries by physicians vs by ChatGPT. And I think it stunned a lot of folks because the ChatGPT-generated notes compared quite favorably to physicians', as rated by patients and physicians.

Whyte: I think that's the [_JAMA Internal Medicine_ piece](#) you're referring to. Let me hear your thoughts on that.

Manrai: Yes, the _JAMA Internal Medicine_ paper. What was amazing is that that narrative sort of got upended overnight. Now we're grappling with that. But to answer your question, many health systems are piloting this now. I think we will see this. Not only is there the ability for these systems to generate responses quickly, to respond to many different aspects of the queries, to respond longer than a physician has time to respond to patient queries, but there's also a simplifying function that is very interesting that AI seems to be very, very good at. This is taking complicated language and jargon and making it understandable for a broader group of individuals. With the digital scribes, we're going to see that as a very common feature of the medical records system.

Whyte: Eric?

Topol: That UC-San Diego study was a bit contrived because it was comparing Reddit volunteer doctors vs ChatGPT. But the actual sense is that when you have these notes

that we've been talking about, that are synthetic with the conversations, and you put them through GPT-4, the machines don't have empathy but they're really good at promoting empathy. So it turns out that it criticizes the doctor for not having listened more or being more sensitive or expressed empathy about this or that. And this is really good coaching. Because we've been squeezed for time so much over the years, the average patient gets interrupted after about 8 seconds and doesn't get to tell their story. So here, I would anticipate that in the years ahead, maybe *not* years, we will be using AI to help coach physicians with these notes that otherwise never would exist, to make them better at being empathetic. It will help in rebuilding that relationship that has been hurt over the years because of not even having face-to-face contact, no less the gift of time. What we're talking about is indirect (the gift of time) but also direct — the surprising capability of these large language models to support better, compassionate communications between two human beings.

AI in Diagnosis and Treatment: The Hope and the Fear

Whyte: I want to turn to what you all talked about, which is where the real power of AI might be: as a key asset in diagnosis. I mentioned before that about two thirds of physicians think that AI could play this role. But as you'd expect, they also have some apprehension about AI in diagnosis and treatment. About a quarter of the doctors in our report were very concerned with AI driving diagnosis and treatment decisions, and another 41% were somewhat concerned. This brings up some very gut-level issues regarding AI. Are doctors concerned about the competition, in the sense that they've spent hundreds of thousands of dollars on medical school and then AI vendors are saying that they can do it better than doctors, so to speak? Ted, what are your thoughts about this in terms of the narrative about how to use AI and the practical aspects? You're in the surgical field, so you don't want to make mistakes.

James: Right.

Whyte: So, what's its real role in diagnosis and treatment?

James: I think the main narrative here is that AI could be used to augment what physicians do. It's not really about replacing clinical judgment or replacing our skills or our intuition, but it can augment all of that. I'll paraphrase something that Eric said: It's not that AI is going to replace physicians. But physicians who use AI will replace physicians who don't use AI.

Whyte: That question has come up: Are you at a competitive disadvantage? Are you going to be at a competitive disadvantage if you don't start to incorporate AI?

James: It's like anything else with any technology. Whether it's the stethoscope or AI, if you're not using a tool that is going to be beneficial for your patients, then yes, you're going to be at a disadvantage. But if you use the tool and you use the tool well, then you'll have the competitive advantage. I do think that AI is going to be seamlessly integrated into office practices and help us with diagnostic capabilities.

Whyte: Raj, what are your thoughts about how AI is going to play a role in diagnosis and treatment? Who fact-checks? That's the question that came up. Who fact-checks? Is it the doctor? Is it the system itself?

Manrai: It's such a critical question. If you told me a year ago that we would have a general-purpose AI model that could take the presentation or case portion of several dozen *New England Journal of Medicine* CPCs, the clinicopathological conferences, and not only get some of them right but do almost stunningly well out of the box without any prompt engineering, without any real tuning, I would have said it was science fiction. Even a year ago, because ChatGPT was end of November 2022. And now we have almost dozens of these studies: ChatGPT, Claude, Llama 2 — a fine-tuned model that is evaluated on some interesting specialty or USMLE set of questions, and they are performing so well and so *consistently* well on these types of exams that we're desensitized to these papers when they come across our desk. I think that is amazing. But you're hitting the nail on the head here when you're asking about who's checking it. What does that evaluation actually mean? What that evaluation means is very impressive; it does feel like science fiction to me still. But how that influences clinical practice or how that gets integrated into a clinician's workflow is a completely separate question. And we have very little evidence. I think we're going to have more. Hopefully, over the next year we'll have some better studies. I think some are underway, but how humans change their diagnostic reasoning process, how they improve, how they don't, how they anchor — maybe they anchor to what the AI is saying in suboptimal ways. What that human-AI interaction looks like and what it leads to in terms of improvements to reasoning, we don't really have strong evidence right now, but I am very optimistic. I think we need to do the studies, though, and where we've gotten to is that we've motivated a lot of these studies, but we haven't really shown yet that AI that can pass USMLE questions, just dropped into residents' hands, is going to improve their care for patients.

AI Can Have 'Hallucinations' but It Can Also Monitor High-Risk Patients Remotely

Whyte: Passing a board exam is very different from seeing a patient. A question came up in the Q&A: How do you address AI hallucinations, which is basically just making up information? There have been some studies that have shown that over time, hallucination increases in some of these models. How do you know? How do you monitor it? How concerned are you about it?

Topol: It's an important area that obviously needs constant work. It goes back to what Ted said. The human in the loop is essential. You don't want to ever trust the output. Over time, it's projected that these will improve the error, the fabrication. I mean, it can be scary to see it. And as you just touched on, John, there's another problem that isn't generally out there, and that's the issue of performance degradation of models over time.

Whyte: You said it better than I did.

Topol: I think [James Zou at Stanford](#) has published on this. It's a common experience. It isn't just that you have to have the sense of being always circumspect, always providing oversight. There were some actions taken by the lead tech companies to put in guardrails, and that led to some different performance. But this is beyond that, so we've got to keep our surveillance up. I was going to comment in response to what Raj was saying. We're getting more and more autonomy for patients, for their own doctorless diagnoses, or at least screening. For example, some years ago, the first deep learning algorithm that was approved by the FDA for consumers was diagnosing atrial fibrillation through a smartwatch. And we figured out how to do that. The problem is, of course, it's marketed to everybody when only certain people need it. But it will say atrial fibrillation or not, and often it will be correct and sometimes not. It always requires a doctor to help sort that out. And

now we have this emerging for urinary tract infection, children's ear infections, skin lesions, and skin cancers. The list is going to keep growing.

Whyte: We have it emerging for blood pressure; Apple has just announced that for their watch. It's going to be the outliers as it is for AFib. But think how that can decrease cardiovascular risk.

Topol: Yes, and I just want to emphasize that we're talking about AI for clinicians, but there's a lot of AI for patients simultaneously, and that will help decompress the crazy lives that clinicians have.

Whyte: Ted, one of the questions that came in asked: Can you give me one good example of how AI is being used right now in clinical practice? And one bad example?

James: I think I probably am biased. I focus a little bit more on the good examples. But there are a number of organizations that are really being pioneers in this area. There's an organization that's using AI to monitor patients once they leave the hospital. I think this is a fantastic thing because, obviously, we're not there 24/7. They'll have a chatbot checking on a patient on a daily basis, either by text or verbally, and they can see how the patient is and track their progress, and alert the clinicians if there's deviation from where they should be on their path to recovery. Some organizations and cancer centers are doing this now after patients are sent home with an oral chemotherapy. They're doing this now for a surgical/postsurgical follow-up. There are organizations now that are identifying high-risk patients, like someone with congestive heart failure, and checking to see if they're developing symptoms. And they're showing that this is actually keeping people out of the emergency department, partly because we can identify when people are getting into trouble early on. So I think there are a lot of very good examples. In terms of the bad examples, going back to my first statement, it's really anything where we try to completely remove the physician from the equation. As Eric and Raj said as well, AI is fantastic, but it's not infallible. It will hallucinate; it will make some mistakes; there are going to be work errors and things of that sort, so it still requires some oversight. Anything without that oversight, I would say, would be a bad example.

The HIPAA Question

Whyte: Raj, one of the questions that came in relates to HIPAA: How does HIPAA apply to these tools?

Manrai: This is a rapidly evolving landscape around data confidentiality, a secure environment where you can have patient data, patient queries, entering into and being used in conjunction with these models. One of the complicating factors here is the variety of models. There's ChatGPT but there's also GPT-4. There's ChatGPT powered by GPT-3.5. There are different models, they're changing, and there are both proprietary and open-source models. So this really is a complicated landscape. There's a real kind of big-bubble effect here, where people might not even realize what is going into it when they enter a query into the ChatGPT interface — where the data goes, where it can be stored, who can use it after that and how that's different from, for example, using the programming interface API for entering the data. There's a lot of complexity. But again, several health systems have moved forward. There's a Microsoft version of this on the Azure Cloud now that is secure for some of their pilots, the studies that are being done on implementing GPT. But assuming that your data is private, assuming that you are not actually entering data into

the training corpus for one of the companies that is creating one of these extremely powerful models — I think it's a hard assumption to make for these interfaces.

Whyte: And that's a whole other conversation about HIPAA. There's a lot of discussion on the Hill about it being time to revise HIPAA. If you use ChatGPT or another service at home that is asking about your medical conditions, that's not a covered entity. But if you're a physician office or health system and you deploy an AI tool to engage in a chatbot, then it's a covered entity and then the HIPAA rules apply.

Will AI Provide a Better Link to Genomics?

Whyte: Before we go on, I want to go back to a question from the audience for you, Ted, about how AI is going to help diagnose cancer early.

James: I think it certainly can, especially if we're going to be using it to track patients who are at high risk, with high-risk behaviors, directed toward high-risk prevention or interventions. It can also go beyond that, beyond early diagnosis. There's an incredible need for using AI to benefit patients who are at higher risk for cancer and have preventive strategies. One of the things that we're planning to pilot at our breast center is using AI to scan patients' records and identify, based on family history or other known factors, a patient's risk, and then directing them appropriately. Surprisingly, a lot of patients may not be aware, and physicians may not be doing the screening routinely. So having an AI system that can monitor for that would be fantastic.

Whyte: Eric, we've talked about that in familial hypercholesterolemia, in terms of looking at labs, looking at triglycerides, talking about apolipoprotein. Haven't we done that in the cardiovascular space too?

Topol: Yes. Defining risk is going to be much more than just labs. Today in *The New England Journal*, nearly 60,000 Icelanders had genomes assessed for all of the genomes on the American College of Medical Genetics pathogenic list. A lot of people were carrying, unknowingly, *BRCA2* variants, and the risk of dying 7 years earlier was tied to that, with sevenfold increase of developing multiple cancers like ovarian, pancreatic, breast, prostate. We don't do this; we don't accept genomics; we just keep ignoring it and the studies keep coming out. There's another study that I think is really important, looking at pancreatic cancer in the whole country of Denmark and the US VA. If you just look at the electronic health record, you could pick up who's your high risk. And so instead of diagnosing people at stage III or IV, we should be finding the people at high risk so they're under tight surveillance.

Whyte: What's holding us back?

Topol: We're resistant to using genomics in general. Once a person is diagnosed with cancer, they'll be screened for a mutation. But that's about the [extent of the] use of genomics in adults today, unfortunately. And then there are these multicancer, early-detection blood tests, the liquid biopsies. They're being done in people who are 50 or older, just like how we do colonoscopy or mammography. That's so dumb, okay? We should be doing these tests in people who are identified as being significantly higher risk. And the same goes across not just cancer, but cardiovascular, neurodegenerative diseases — anywhere we have an intervention to keep them from getting this bad diagnosis or so we can intervene at the earliest possible time. AI is multimodal now, with transformer models that can bring in all this data from a person's electronic health record,

unstructured text, genomics, their scans, their pathology if they have it, their environmental exposures, their gut microbiome. So we can define risk if we're willing to go after it, but we haven't been.

Whyte: So how do we get there, Raj?

Manrai: We'll solve the problem right now. The parallel with genomics is a very, very interesting and important one. There are a lot of lessons from the way genomics has evolved. I think what Eric is getting at is that maybe it has not penetrated enough into aspects of practice, that we need a comprehensive picture of the patient. Genomics is one layer, labs are one layer; imaging is another; the environment, social determinants of health; where patients live, work and play. All of those things are important. Thoughtful clinicians know this and they use that information when they're reasoning, but they have access to only snapshots, only little bits of that information in that very limited patient encounter. Let me take both an aligned and somewhat not aligned position. I am fully aligned with the vision. I think we need the comprehensive picture of the patient. I think we need to identify those patients early on who are likely to be at risk. Genomics offers one path to that and AI offers a very interesting and complementary path to that. The other part of me will say that we should be doing that, but there are false positives. There are risks of trying to identify disease in a population with a new AI test or with a genomics screening panel — that maybe we have validated *BRCA* variants where we have a rigorous understanding of penetrance, which is the probability that an individual with one of those variants goes on to have the disease, which Eric was referencing. But for every one of those, we have thousands and thousands of other genetic variants where we have no quantification of the risk that's actually attached to that particular genetic variant. We have pathogenicity reversals where variants that we thought were pathogenic are no longer considered pathogenic or disease-causing. When I was in grad school, I published a paper in *The New England Journal of Medicine* focused on inherited heart disease, hypertrophic cardiomyopathy. We found that not only were there variants that were misclassified as disease-causing that were not actually disease-causing, but that these occurred disproportionately in African Americans. And this is because of a dearth of controls in some of those original sequence studies and original clinical studies that link those variants of disease. I think there are a lot of parallels with AI. There's a lot of amazing AI, there's a lot of hype, and there's a lot of bad AI too. There's a lot of opportunity for not quantifying risk and not understanding what the implications are of broadly screening a population that's at low risk for disease, even if you have a seemingly good model.

Will AI Increase Malpractice Risk?

Whyte: I want to turn to a different type of risk, malpractice risk, for which there has already been a bunch of questions. We addressed it in our Medscape survey — whether AI could increase the risk for medical practice that leads to a lawsuit or whether it could decrease that risk; 43% of physicians think AI will increase malpractice risk but 29% think it might reduce that risk. Raj, what are the malpractice implications? I'm going to include a question from the audience: How does liability apply if AI makes a diagnosis or recommends a plan? Who's liable if it's wrong? And the second part of it, which I hadn't thought of: If the physician disagrees but the AI was right and you ignored it, how does that change liability?

Manrai: The short answer to the question is that we are rapidly trying to figure this out as a society right now. I don't think there's clarity on this. Those are excellent questions, and I think the audience member who asked them is pointing out that there's risk on both sides:

Either you're using your anchoring to a false diagnosis or a wrong diagnosis provided by the AI model.

Whyte: And that's usually what we focus on. To be fair, we don't usually think of the other side; I was right.

Manrai: If you were to ask leaders of health systems how long it will be until it might be considered malpractice, or suboptimal care, to not have AI as a second opinion — to look over your shoulder or vet what you're saying — I think you might get some interesting answers. I'm neither a lawyer nor a physician, so I should disclose that. I'm going to defer to my two physician colleagues to give you a more nuanced answer about current ways of thinking about malpractice. But my sense is that we are still trying to figure this out. I don't think there's any clarity right now on using an AI tool. If it's saying X, Y, Z and you depart from that, whether you're liable or the company's liable… I think FDA and Congress are trying to figure out exactly how to even regulate these tools, how to build these tools. There was a recent [executive order](#) from President Biden on what is an allowable amount of computation in a given center or for a given model. This space is rapidly evolving and there's a lot of uncertainty around those questions right now. But I'll defer to Eric.

Whyte: Let's hear from Eric and Ted. What's your guidance for that person who asks what happens if the AI was right and you didn't follow it?

James: We have these computer physician order entries. So you try to order an antibiotic and it says that it's a reaction, based on something else they have; you have the option to override it. So we've kind of lived in that world already, where you could potentially make a decision that goes against some automated thing, and obviously there's going to be liability there. But on the other hand, I'm hopeful that with the use of AI, helping us to broaden our differential diagnosis, helping us with rare presentations, helping us to align well with best practices and guidelines, we may see greater patient safety and fewer errors. I think the net total will be more toward reduced liability and reduced malpractice. Eric, what do you think?

Topol: I think the big issue, that physicians don't like to admit, is that there's a serious problem with diagnostic medical errors. We know from various studies, summed up by the [National Academy of Medicine](#), that we're talking about 12 million a year of serious medical diagnostic errors. Recently, [Johns Hopkins](#) had estimates published in *BMJ* that about 800,000 Americans either die or are severely disabled each year from medical diagnostic errors. So when do you think it can be made worse? Most of us don't think they ever make any errors. I understand that. But we have a tool here that can help. Ultimately, the physician is always going to be responsible. But it isn't as much on the treatment side. Often, the problem lies on the diagnostic side. This hybrid, where the AI helps here, we haven't yet seen it yet, John; but eventually we're going to see the corpus of medical literature real-time as we're seeing patients. Who can keep up with all of that? I mean, I try, but it is very hard. So if you have all the data that is integrated about a person, the chance of you helping to formulate a correct diagnosis should be enhanced. We're not there yet. There's no generative AI with the whole medical knowledge base that's available yet. That's going to require a lot of fine-tuning and regulation, but we will get there in the months ahead.

Whyte: Do you expect the government to take a bigger role in the regulation of AI and healthcare? Does anyone think that?

Topol: It depends on which part. It's not going to have a regulation of using administrative tools, like we talked about, but it likely will [be involved] in other aspects. There are lots of different use cases here; it's not a simple answer.

Is AI the Answer to Burnout or an Exacerbation of It?

Whyte: Ted, the question came up about burnout. Do you think AI is going to help alleviate physician burnout? A lot of things had been promised that were going to reduce burnout, and they haven't. But is there a concern that AI enables you to see more patients a day, it enables you to engage with more dialogue, but that it potentially could increase burnout?

James: Yes, for either one. AI is simply a tool; what really matters is how that tool is implemented. If it's implemented in a way to drive revenue and to optimize billing, it probably is going to exacerbate the burnout issues that we have right now. On the contrary, if it's implemented in a way that really helps to streamline workflows, or to provide ease of practice and efficiency in practice, then it could reduce burnout. But it's really going to be on a case-by-case basis in terms of how the tool is implemented and who's driving that implementation. It's one of the reasons why, as physicians, we need to be at the forefront of this — to make sure as much as possible that this will be implemented in a way that improves our ability to take care of our patients and to do so in settings that don't burn out our colleagues.

What AI Could Mean for Other Clinicians and Researchers

Whyte: One of the questions that came up, Raj, relates to nurse practitioners and physician assistants. Is the use of AI in these professions different than it is for physicians?

Manrai: It's a great question. In many ways there are more similarities than there are differences between how those tools will evolve. And I suspect that, with well-defined knowledge bases that exist underneath these language models that they can draw on, many of those tools will be the exact same tools that both physicians and nurse practitioners and other providers are using. So, I could see the landscape evolving such that there are well-defined and specific tools. But I don't think the broad contours of the problem are really that different in the ways that nurses vs physicians will use these tools.

Whyte: Eric, do you think there's any difference in how it's going to be used?

Topol: I don't think so. Eventually, when these things are built properly, they'll be broadly used. I don't know that there are going to be significant differences between professionals.

Manrai: Eric mentioned "multimodal." There was a big announcement from OpenAI, the creators of the GPT model — GPT-4/ChatGPT. They announced about a month and a half ago that the chat GPT model now has vision, so you can put text in there. We all know that you can interact with it as a chatbot, but now you can take pictures with it. So everything from my 6-year-old daughter and 4-year-old daughter — you can take a picture of their homework and talk, and they can look at it, and you can talk to the model and ask it to rephrase the homework, reframe it with their favorite characters from a book. It really is amazing, amazing out of the box, for some of those types of applications. And there are other interesting applications where you can feed in biological medical images and interact with those images as well. We've seen some evidence, both positive and negative, for some of those applications. But I think it occurs to me that multimodality might be an

interesting axis in which nurses and physicians might have similar but also different uses. If you think about the workflow, how nurses are interacting with patients and in the hospital, there are multimodal handheld or other AI devices that might be extremely useful for some of their tasks as compared to what physicians are doing. And then I also just had a silly idea. Eric has this amazing and well-deserved Twitter (or X) following. As I'm sure many of the listeners know, he highlights papers — the key aspects of papers. I think there's sort of a window into what Eric is looking at. Many doctors I know, many researchers I know, follow Eric's Twitter profile just to learn how he's reading the papers and keeping up with the literature. And now that we have multimodal AI, I bet that is an amazing source of highlights, to train the AI model to look at how a cardiologist sees the literature and keeps up with the literature. So that is a little silly thought that just occurred to me because you're mentioning multimodal AI.

Topol: Thanks. We saw over 200,000 posts on Twitter by pathologists; they put the histo slide and their interpretation. And that was used to develop an incredible tool to interpret pathology slides. Who would have guessed that Twitter could do good? For pathology. There's one thing that I think we should underscore: It's our imagination only that's holding us back, because what we're seeing now with this tool — a powerful tool and a two-edged sword — is that it's doing things that we would not have anticipated. The recent publication we had with the [retina photos in *Nature*](#), which showed how with just these million-plus images of the retina you can predict Parkinson's, Alzheimer's, heart failure, heart attacks…

Whyte: You've always liked the eye, Dr Topol; you focus on it in your books too, even as a cardiologist!

Topol: The eye is like the gateway to the body, to me.

Whyte: To the soul.

Manrai: How is this model working? How can it predict all those things? What is it? What is it seeing in the eyes?

The Explainability and Ethics of AI Are Still Elusive

Topol: This is a part of the problem that John hasn't asked yet. It's about explainability. So before we got to these transformer models, we already had a problem with explainability. Now we've got that multiplied. Nobody understands how these models work as well as they do when they do work well, or why they hallucinate. This is a problem we have right now. We haven't really gotten to the nuts of how these are working so well. That's why you said that the winners of the Lasker Prize, the DeepMind crew — Demis Hassabis and John Jumper — should've had an asterisk for AlphaFold, because they don't even know how it works. It cracks the 3D structure for 200 million proteins. We have to acknowledge that this is an early time for applying these tools at scale, and we still don't understand them well.

Whyte: It's the explainability, but it's also the ethics. And that has come up. I wanted to ask Ted about the ethics of using AI in the diagnosis and management. How do you explain it to patients? Do you disclose it? What's the duty of physicians to talk about the use of AI in patients' evaluation and management?

James: That's a fantastic question. You had asked me before to give you examples of bad AI, and I gave you one, but I'll give you another one. I think utilizing AI in patient care without full transparency would be another example of bad AI. So whether you're using it

to help you generate the diagnosis or to personalize your treatment plan, or to connect with patients, to monitor them to identify for high risk, I think it's important for patients to realize that and for physicians to be fully transparent with patients about how we're utilizing the AI, why we're using it, what we're going to do with their information, and how this is going to impact their care.

Topol: Ted knows that when he's operating, anesthetics are given, and we have no idea how they work. And imminently, psychedelics are going to get approved by the FDA. We don't know how they work. So the question is, if we have AI that's doing great things, validated in rigorous, large-scale studies, randomized trials, and we don't fully explain it, do we hold the AI to different standards than we do with the medical interventions of today? That's an unanswered question.

Whyte: I asked you that the other day.

Topol: Well, I think if we accept things without proof of how they work, which we do widely in medicine…

Whyte: Yes.

Topol: Are we going to hold machines to a different standard? That hasn't been grappled with yet. There's compelling data: Lots of research in AI in medicine is largely absent. So, we'll see.

How to Get Your Practice Ready for AI?

Whyte: We titled this webinar, "Is Your Practice Ready for AI?" So I want to go around and get one or two pieces of advice for viewers. How do you get your practice ready for AI? We've all kind of acknowledged that AI is here. Those who don't use it are going to be at a competitive disadvantage, ultimately. So how do you get your practice ready? Raj, let's start with you.

Manrai: One of the major themes of this conversation is what we've called human in the loop, or doctor-, collaborator-, human-AI partnership. It's referencing the importance of the physician overseeing what the AI is producing, knowing about some of the blind spots of the AI, knowing its tendencies, knowing where it works well, where it doesn't work well. So in addition to everything that we would normally approach clinical practice with, we're trying to understand safety, efficacy, how well this works in different populations. Large language models are just one type; this also applies to others. One of the best ways that doctors can get ready for this in practice is just to use, play with, and test the model a lot. What I found in speaking with physicians here in Boston, some of my collaborators, my friends and family, and when I travel outside of the city and give talks, there are amazing bubble effects where we have some doctors who are using these tools and almost developing relationships with GPT-4 or with Llama 2 or Claude, and others who used it once. They weren't that impressed, and then they stopped using it. We've seen the numbers: 100 million users. I have no reason to doubt it. But I think there are very strong bubble effects where some people are using this a lot and other people are barely using it. You start to learn where it trips up, where it makes mistakes, what its error modes are, what exactly that confabulation or that hallucination looks like, and you start to develop a feel for it.

Just to bring in an analogy which a lot of people like and a lot of people don't like but which I think is I think is net-useful: If you think about something like full self-driving of a Tesla, there is a lot of transparency in Tesla, for example. There's a lot of transparency in the display of the way that model sees the world. So you see what other cars it sees, you see what cones it sees on the road, you see what signs it sees, what it doesn't see. You see the stop signs and then you see its path, right? You see which direction it's going to go in. Over the past couple of years this has been rolled out and now is very effective. I think we're still missing some conclusive data on its exact safety profile with respect to human drivers, but it's very effective. And there's a lot of trust built in by the humans who are using these cars because they can grab the wheel. They see its state-of-the-world model and they know how it sees things and what it's going to do, but they know where it fails. So they know as it's getting better, as software updates are going out. It makes crazy mistakes sometimes, but it makes way fewer of those now. But when it makes those wild mistakes, you sort of get a feel for it as you're using it, as you're interacting with it. And you know its warts, you know its problems, and you can be especially attentive while it's on autopilot — ready to drive, ready to grab the wheel. I think there are a lot of lessons there for how we think about deploying AI in clinical practice. But for physicians, it starts with having a deep familiarity with the problems with error modes of the models.

Whyte: We talk a lot about it, but it's actually a very small percentage of the US that has actually tried it.

Manrai: Using it after 5 hours is very different from using it after 5 minutes.

Whyte: I'm not sure how I feel about the self-driving cars.

Manrai: It's polarizing.

James: In terms of getting your practice ready, I think the number-one thing is for physicians to be very curious about AI. I don't think we should try to be overly skeptical. I don't think we should be afraid of this. I think we should embrace it because it's a wave of change. And I think those who embrace and experiment and learn, as Raj is saying, will be able to develop capabilities and really leverage the AI for everything that it's worth. And those who don't, quite honestly, will be left behind because AI is not going away. It's going to continue to get better. I do believe it will become a mainstay in practice. So be very curious, experiment, and also think about it from your patients' perspective. How can we use AI to enhance the patient experience, to make the patient journey better, to have better patient engagement? I think practices and organizations that can do that will be the leaders in the future.

Whyte: Eric, you have the final word.

Topol: I would really encourage every clinician to read the book by Peter Lee, Carey Goldberg, and Isaac Kohane, about GPT-4 and the revolution in AI. If they haven't read it, a lot of things that we've talked about today will get reinforced. And obviously, getting familiarity is important. I recommend not ChatGPT but GPT-4 through Microsoft, getting Microsoft Edge, Bing, and go into creative and just work with that, because that's state of the art. It's free, and you'll have a better experience than you will get through ChatGPT, and it's multimodal. You can put in things like scans and whatnot, so everyone should get familiar. When I first got that book from Peter, I stayed up. It's a quick read, but I couldn't put it down. Hopefully everybody will get a good primer for where this is headed.

Whyte: Eric, Raj, Ted, I want to thank you for participating tonight. I want to thank all of our viewers. This was our last Medscape Masters for the year, but we'll be back in 2024 and I hope to see all of you then. Thanks for participating, everyone.