# (no title)

25 MIN · YOUTUBE · HTTPS://WWW.YOUTUBE.COM/WATCH?V=W8Z2O0ZV2SA
https://www.youtube.com/watch?v=W8z2o0zV2SA

SUMMARY

*Dr. AB Abernathy discusses the rapid advancements in AI technology and its implications for healthcare, emphasizing the gap between AI capabilities and their practical application in clinical settings. The conversation highlights the challenges of integrating AI tools into medical workflows, the need for better understanding and training on these technologies, and the importance of generating evidence for effective implementation.*

- *AI capabilities are advancing rapidly, but their adoption in clinical practice is lagging.*
- *Physicians often use AI tools sporadically and lack training on how to effectively integrate them into their workflows.*
- *There is uncertainty about the legitimacy of using AI for critical tasks, such as interpreting medical tests or communicating with patients.*
- *The healthcare system struggles to keep pace with technological advancements, leading to a disconnect between AI potential and real-world application.*
- *Evidence generation for AI implementation in healthcare is lacking, necessitating a focus on best practices and optimization.*
- *Regulatory frameworks for AI in medicine are still evolving, requiring flexible governance to adapt to new technologies.*
- *The integration of AI may reduce the demand for certain types of clinicians but could also enhance the need for human interaction in patient care.*
- *Ongoing discussions are essential to navigate the complexities of AI in healthcare and to develop effective strategies for its implementation.*

Welcome back to the Stanford Healthcare AI podcast and we're so excited to be joined by Dr. AB Abernathy who really needs no introduction but a physician scientist formerly ran Verily Flat Iron and roles at the FDA um and is now running Highlander Health which is promoting companies that generate new evidence uh both from uh uh investment and uh philanthropy perspective uh and we have a ton of topics uh to to run through today and all the advancements.

So welcome Amy. Well, it's awesome to be here with you. Thank you. Uh so the the the first topic we wanted to jump into uh and I'll pull up some data uh is all of these new advancements. And so Matt, help help ground us into all these new things we're we're seeing here from from the new AI companies and models. Yeah, it's it's really interesting. I know we talk a lot about the acceleration of the capabilities of of these models. Um, but it's almost I don't want to obviously we keep uh pretty close tabs on it, but in general I feel like, you know, we see results like what you're seeing on the screen, which again we can argue whether this is the way that you would test quote unquote intelligence, but nonetheless, we see this repeatedly that these benchmarks that we've put out there, whether we're trying to make them harder or more focused on certain areas, they keep getting saturated,

right? But I also feel like um uh it's it's not really reaching the public consciousness maybe as much or at least the the physician community. I'm you know folks maybe still dabbled in you know Chad GBT 6 months ago or maybe they're using it occasionally for a quick question answering but are we really like taking advantage of what I'm starting to see as an overhang of capability versus uh what they're being used for? I I don't know like it I guess my point is would we even know AGI if it if it happened and would anyone even talk about it at this point right because it just seems like the advancements are so uh fast.

I mean it's interesting and I'll be curious what the two of you think as well if I let's start in the beginning Matt. Um and your point is that the you know capability sets of of the models that we have in our tool suite right now are largely way more powerful than probably how we're using them as clinicians uh today or even understood understand that they could be used. And you know a lot this is about adoption curve right you know figuring out how do you put a solution into your day-to-day workflow. How do you understand a combination? I'm going to leverage this solution and I'm going to also be able to communicate and I'm going to come back to the clinical context in just a second. Communicate what I've learned in a way that now helps for example the person sitting in front of me know what to do next.

And you know, you were saying, Matt, like people are largely just dabbling and, you know, trying things out, but a lot of what is going to be needed for uptake is that dabbling. Um, I I personally, you know, some days I use Gemini, some days I I use Chat GBT, like it kind of changes around on the day for no good reason to be honest. Um, and I know I'm not pulling um the greatest capabilities out. And part of it is because frankly I'm still trying to figure out which use cases are most productive. I'm still practicing how to prompt. Um, frank frankly in in that particular scenario, my son is way better at teaching me how to prompt than anything I read I read in the medical literature or online. So my 24 year old is more effective than um you know papers about prompting in medicine. So I I need to learn how to prompt. I'm trying to understand also within the portfolio of my work what is legitimate.

So for example I am trying to figure out if I was working on an editorial over the last three days. Um is it legitimate to ask LLMs to help me edit? Is it legitimate to ask them to help me think? Is it that they're a brainstorming partner? Is it, you know, kind of all of the above? Or should I just tell the LLM, go for it, write my editorial, I'll see you in a day or two. Um, and and so in thinking about how to legitimately embed LLM into my work, there there isn't a training manual for that either.

And then you take something that's way higher risk than writing my editorial, which is the care of a patient. Yesterday, my problem was my mother uh had a um series of radiology tests, sent me the tests, and then wanted me to tell her what was going to happen next. And and I was trying to decide like how legitimate is it for me to ask ChachiT to to to help better communicate all this information to my mother and where's the liability? How do I know I'm communicating the right information? I it was a radiology test that I actually didn't know how to interpret. So I didn't have a way of being sure the LLM was gonna um be able to do its job. So the first part of this is really my saying to you. I think we're all really learning how to incorporate LLM into our work in our daily life, how to master them and what's legitimate. Then you jump to this other question that you asked which is would we know AGI if we saw it? And given the fact that we're kind of bumbling at the moment, I suspect we

probably wouldn't know AGI if we saw it. Um, even if we were really bullish AGI is coming tomorrow. Um, but I'm going to stop there with my saliloquy and ask the two of you, would we know AGI if we saw it?

it it's hard to keep up and and I think this is something and this is part of the reason I think we're doing this is we struggle to keep up with the current advances and it's more or less Matt and I's full-time jobs in in the work that we're doing focusing on AI and we're struggling to keep up testing all these all these new pieces and I think Amy the point that you brought up that's so interesting and I'll pull up this one other chart is uh the consumer adop option is crazy of these tools, right?

Everyone is using them at home. You know, you're a physician and you're now saying, "Wait, hey, should I be using this to interpret radiology results?" Because it will help explain this to my mom, but you're uncertain, right, of the capabilities. And these are just the real issues that I don't think we are talking enough about as a community in medicine because we kind of pretend it's not happening. Many of the health systems kind of pretend, well, we're not going to give people access to these tools. People aren't going to experiment. But it's happening whether we like it or not on just how many people are using these tools now. And so uh no, I think to to your question, I I don't think uh we we'll know AGI when we see it in healthcare. Yeah. I there's this piece that feels so 2000 to me.

Well, maybe it's 1995, but I, you know, I'm an oncologist by background. And there just became this period of time when every patient that walked into the clinic had a stack of paper. They printed off probably on a dot matrix, but you know, sort of on their printer of things that they had pulled off the internet. And there was some combination of help me make sense of all of this from the patient and shame on you for not knowing all of this um from the patient or maybe I was sort of shame on myself because I was like thinking to myself, holy cow, I don't know what this all means. And um the that you know as clinicians we felt a little baffled about what to do. You could just say don't listen to the internet. Well, that seemed like kind of a silly uh point of view. and for us as clinicians um to ignore that our patients are incorporating LLMs into their daytoday life, let alone into their clinical questions. Sort of seems silly. We could go all the way on the other side and say use LLMs for everything, but we know that LLMs are fallible and so that's probably um you know m moving too far in a different direction. So guiding patients thoughtfully down this path is the new 2025 bless task. It's really been with us for a couple years now and um probably should look back to how did we adopt quickly in the late 1990s and what were some of the lessons learned?

Yeah, I I really like that analogy because um I can't think of a clinical day I've had in the last 10 years where I didn't use the internet at least one time. Right. But that right but in the '90s that you've been like why would I use the internet? I just, you know, um, but what's what this progression I'm sort of feeling and I I don't know if I've articulated this quite fully formed thought yet, but I feel like it, you know, computers, let's just say, were like, okay, they they offloaded our need to do computation, right? And then with internet, I feel like they offloaded our need to remember stuff at scale. And then with the social media, it's kind of offloaded like our attention is now gone. And I feel like this frontier is like our cognition is now being potentially threatened. Do you know what I mean? Like it feels like there's this slow progression. Then of course we can get into robotics which I don't know if we have enough context around but just for me like how much longer will it be till it's

normal for just like I would use the internet today. Uh would I use LM just for everything just interacting with them. Hey this is what I'm thinking about doing for this treatment. uh this is the test I'm thinking about like just having these conversations or frankly even agents and abstracting that even further and say hey can you go run and u do an analysis on this on this chart before I have a convers it's it's really fascinating but I think that the part of this is the daily use like you said the dabbling to get the familiarity and understanding the capabilities and on the other side the patients are kind of pushing us I see it already too it used to be like you know your Google search isn't a replacement for my medical degree kind of thing because right I mean there's a lot of sources you could be like, you know, where did you go? But now it's like, well, which model did you use and how am I supposed to, you know, how am I supposed to have all the understanding of where the capabilities of each individual model are? You know, we've done some work to for these benchmarks that we're just trying to barely get our hands around. And it's we know we can't use med QA like as law in terms of what they can do. But yet we keep finding in these head-to-head comparisons physicians with AI physicians alone AI alone that AI just tends to do better than AI plus physicians which is kind of calls into question even if I'm using these tools am I am I still uh performing as well as I need that this is some some work from from Google um that was published in nature around their AI AMIE system um which you know admittedly is somewhat more geared towards um you know medical tasks but but look at these lines like these trend lines like even the clinician with that tool underperforms um and and this is not just an isolated result like we see this time and time again in fact to the point where when I used to be asked will AI replace visions I'd be like of course not it's always going to be AI plus physicians and we just keep seeing these kinds of results here's another one 01 preview by itself outperforms physicians using these different tools So, I don't know the answer. And part of it, I think, is do we have the best practices of how we interact with these models? We're we're kind of fumbling around the dark, right, when we're using them. But still, I I don't know what to what to sort of um what to advise folks at this point.

Well, so I'm going to go back to something you just said about sort of the the sequential offloading, right? Um so so I'll start there first which is yeah we're at this place of as you described it offloading cognition to LLMs. Um and and what that is at least you know if you sort of kind of look through the progression you talked about it was sort of computation at scale it was information finding at at at scale.

So now it's really integration of myriad very duse potential data points now to be able to have a moment of inference or decision, right? Like so that's where where we are and and what we start to offload to LLMs is the ability to reach across all of those potential sources of information and make sense of them um in the moment. So kind of that's where we are, but we're not offloading social interaction.

You know, like we kind of talk about, for example, LLMs are kinder and write nicer messages, but at least I sincerely believe some of that is that a as we've had a number of new technologies come online, including, as you mentioned, social media and offloading of our attention, our ability to deal with that overload as humans has gotten caught up in our ability to be empathetic. and emote and that LLMs to help offload some of that demand so we can get back to being more human again seems very important.

The second thing that I would say as it relates to that is that it is not writing nice emails is not the same as being

able to be a confident leader of people of setting vision and course in a way that everybody wants to move towards a combined future. of being able to help a scared patient and family figure out how to navigate a really tough decision that likely incorporates a series of facts for the patient that would not necessarily be embedded facts if you just were were to look at the usual data set. And so, you know, I I suspect that there's some offloading of the of some of the demand of emoting and therefore the opening up of a new frontier that looks much more like LLM's having taken care of some of our um need to push cognitive overload to individual data points and pull some of that cognitive work to the part of the work that's distinctly human. So that's kind of my my and and maybe that's my, you know, um glasses half full point of the world, but that's where I hope to to see us go. And then the other thing that you brought up um and and and the figures highlight is at least right now it specific discriminate tasks, right? differential diagnosis, pulling a whole bunch of facts together and making a medical decision about what to do next. LLMs look better on their own than LLM plus physicians. But just like anything else, probably that's some version of physicians still not you knowing how to use LLM as you had as you said, Matt.

And you know, I I I think about one of the things that we're trained to do as doctors. Um we are trained as doctors and we go through internship right we as as interns um and I don't know what kind of clinicians you're trained as so I think actually both of you are radiologists no internal medicine and radiology is that right uh math math radiology I went into the startup world oh okay well there we go so um you know internal medicine oncology and and when when we basically train internal medicine residents. Part of internship is teaching interns to go from sequential tasks to that moment in time of having the gut instinct of sick versus not sick. At that point, clinicians have hit a new plateau of how they integrate facts and apply it to individual patients.

And that instinctive plateau basically is embedded with many parts of our learning and thinking including our own biases that we have not built LLMs into hitting that plateau. So I would suspect that clinicians pull LM down a little bit because those two things are working at odds a little bit for for some period of time until we learn as clinicians how to do the structured tasks like what we teach interns that transition to integrated thinking and fuzzy thinking to then that transition to fuzzy thinking plus LLM and we I don't know how we you know hope at Stanford you're thinking about how do we train this Um, it certainly hasn't been something I've been talking about yet. We're we're we're we're talking about it. I don't know that we or anyone has all the answers here.

Uh, I'm, you know, the the recent things that we see, you know, over the past, you know, month, you see people like Bill Gates talking about how we won't need doctors in 10 years. Folks at Deep Mind, Eric, we see all these very, very sensational uh, sensational headlines. I agree with you. There's so many components of the work of the human connection that that won't go away kind of despite what these headlines are. But but I'm curious to just touch on these few parts and kind of Amy, you have such a unique viewpoint of being an oncologist, having led at these technology companies, having led at the FDA, like and now you're really focused on like evidence generation and what do we need to show?

And so, you know, we pull up these charts, we pull up these different studies, the evidence seems to be mounting, but what what evidence do we need to start actually diffusing this into into medicine? Because as we

look at it or as I look at it, it seems like we're just getting this divorce between where the technology is capable of certainly what we're training people to do, but then what is actually happening in our health systems? Like what's actually different today? Not that much.

And so like how do we what is the evidence that we need to start to change it to start to make it look different and you know from a few of your hats and especially like how does the regulation tie or not tie into this adoption? Yeah, I mean so many important questions built into one. um Justin and and and and frankly in my mind made worse by sort of the current state of overwhelm of the system itself. Right? So the healthcare delivery system itself is is not set up to you know easily adopt and integrate new capabilities. Partly it's just kind of swallowing water of what needs to happen. Um you know my first observation is we're continuing to look for details and facts around LLM performance but we have fewer details and facts about LLM implementation. So the evidence side around what does implementation look like? How does implementation happen?

Well, right and you know calling it best practices I think is an underwhelming definition because um it it it should be around evidence-based optimization and and and kind of getting to a a fine-tuned understanding. And just think about how much detail we have about the embedding of algorithms of all different types into everything from search engines to all these things that we use in our consumer lives. And that comes from many many different types of testing, causal inference, ABA, like the list goes on. And figuring out that same mechanism to embed and adopt really is an implementation science that that that you know we we really need to be talking about. The other part of that implementation is figuring out implementation plus performance, right?

Because as you implement LLM into clinical delivery settings, the clinical delivery setting itself starts to change, right? And so there now are a series of of ongoing underlying changes including change in the conduct of care, change in the underlying data that is being generated that now feeds subsequent action. And we need to be able to to not only test implementation and study best best ways to implement, we actually need to study how that actually looks across time in the myriad different settings where we're implementing. And I think that's a huge part of of of the work. Um, also kind of embedded in your question, Justin, was, you know, LMS and does AI just, you know, totally take away our need for doctors or clinicians? And and I think sort of like we need to define clinicians very broadly, right? And it probably reduces our demand signal for certain kinds of clinicians, but if we go back to my point before, we've s sort of offloaded some of the cognition, but we've now um increased the space for humanity. We're going to actually need much more of humans in the loop to interpret and make sense for for patients and humanity and sort of point the way. I I still think that's true. which may need um a a differing load balance of the different kinds of individuals and capabilities across the system um to get uh that work done. And then the last thing you asked is sort of from the regulatory frame. I think for the regulatory frame there's a lot of struggle that's that's happening to figure out you know what is the best regulatory approach. You know there's most things that are clearly regulated. So when it's software as a medical device and we're thinking about an AI algorithm to read mammograms and you know we're we're like it's very clear what what we need to do and how we need to think about regulating those devices.

If we look all the way on the other side and we look at LLMs to support the practice of medicine and clinical

decision support, it becomes less clear exactly what the regulations look like. And all of this sort of right now sort of presposes a fixed world where you have to have a predetermined change control plan or something else to, you know, kind of actually to set a set of par parameters up front that may or may not really actually make that much sense or be workable. So I I I I personally believe that we're going to see um the need for more flexible governance systems even both from the standpoint of self-governance from industry and a and and also some of that set through reg FDA in order to set the path of how we go forward. Um, I know we're at time, so maybe like the next time we talk, I kind of give you some ideas of what I've been thinking um along this way and and and we can bat those around.

Would would absolutely love that. And you know, I know we've had some discussions before on this topic, but how do we generate the evidence? How do we think about the monitoring? How do we start to share how we govern these things? I couldn't couldn't agree more. Uh Matt Matt, we'll we'll give you the the last word here before we wrap. Uh I'm uh I'm extremely bullish on the idea that um we will come together and come up with a new category entirely for both implementation but I think for the for the regulatory aspect of this and and how we use it. So um I'll I'll increasingly keep an eye on where the capabilities are, but my my gut tells me that um we're going to continue to have this overhang for the foreseeable future. Yeah. And we're going to have to figure out how to navigate through it.

So, thank you for inviting conversations that at least, you know, start to get to the surface of what we got to figure out. Well, amazing. Thank you so much. Thank you. Thanks.