Thrivecast Episode 68: Beyond Note Writing: The Evolving Role of AI in Medicine

Trish Kritek: [00:00:00] Welcome to another episode of the University of Washington Thrivecast, the podcast designed to help School of Medicine faculty thrive. I'm Trish Kritek, and today we're joined by Dr. Jake Doll. Dr. Doll is an associate professor in the Department of Medicine, Division of Cardiology, and he is the UW Medicine Medical Director of Learning.

And I invited you here, Jake, under the auspices of that second title, which is really one to think about and help us support bringing education to our clinical teams and helping people grow in their clinical roles. And as part of that, I- you and I have had a bunch of conversations of the role of AI in clinical spaces, so I thought that that would be a great topic, super hot topic, to talk about more with our faculty.

So, thank you for joining me today.

Jake Doll: It's great to be here. Thank you.

Trish Kritek: So when I talk to faculty and we talk about AI in our clinical spaces, the first thing that people [00:01:00] go to is helping them in writing their notes. So we have Abridge now here at UW Medicine, and that's an AI-supported tool to kind of scribe during clinical visits.

So I'm curious, though, in kind of the work you're doing and the way you're thinking about it, what else is there in this kind of space that might be helpful to clinicians as they take care of patients and families?

Jake Doll: There's a lot of stuff, and it's moving really fast. And you've identified some early tools that are now being integrated into practice and sometimes integrated directly into the electronic medical record.

But you can also use these AI tools or large language models we may call them, the more general forms of those, to do some interesting and potentially useful things to improve your electronic work life. I would summarize initially the tools that are available right now in, in two main categories. One is really a knowledge synthesis role, and [00:02:00] this would be tools that take the massive amounts of clinical data that we use every day and try to summarize it and condense it down into something that's more usable.

One specific example of this is, I work at the VA Hospital, and often my veteran patients are getting care at outside hospitals, and we do have access through the electronic, health record to many of those documentation sources. But we're talking about hundreds and hundreds of PDF pages, and it is actually possible to grab all that, put it

into a secure large language model, and then say- Tell me what cardiac procedures that patient has had in the last 10 years.

And the tools are actually pretty good at doing that. You can imagine there's other big data sources that we might want to summarize. A hospital course, like none of us really enjoy writing discharge summaries, and there are tools available that can really assist that. The other main kind of documentation support piece is what you've already mentioned, which is kind of broadly called [00:03:00] ambient AI or an AI scribe, and these are tools that will essentially listen in on a patient encounter, summarize everything that happened, and generate kind of a electronic medical record-ready note.

It requires you to review and edit and make sure it's accurate and include the things that were not included and edit out the things that were wrong. But clinical trials have shown that it does reduce documentation, time and, in one study at least, reduced clinician self-reported burnout. So I think there's great promise in those areas, but obviously it does require some education and training to make sure they work well.

Trish Kritek: So, so much there that I want to kind of take different chunks of what you talked about. So the first one that you talked about was, like, taking all this stuff, I like to think of it as what we used to think of as the packet that came from the outside hospital when somebody appeared, and we admitted them into our hospital, but it could be in an outpatient setting as well, and distilling it down or answering a question for us, like [00:04:00] you, you want the cath lab and the echo reports, and I want to know what their PFTs were and whether or not they had a CT scan of their chest.

So we each have our things, and I bet there's clinicians across our system who have those, stuff they'd like to discern from, like, 73 pages of nursing notes and vital signs that take a lot of time. Does that work when you put all of that in there? And, and what are the risks in doing that, as well as the potential benefits?

because I think the benefit seems pretty obvious. Like, somebody else is reading through all of this or something else is reading through all this and telling me what I really want to know. But tell me a little bit more nuanced about that use.

Jake Doll: Sure. In my personal use of it, the short answer is yes, it does work in the sense that it can pull out the information that you need, but there are lots of risks.

Number one, it's got to be secure, and most large medical systems now have a large language model tool that's actually behind the firewall that is secure. You're not releasing that patient information out into the broader internet, though [00:05:00] the models are trained on the broader internet, and so you get the benefit of having a model that has looked at, I can't even attempt to quantify the amount of data, and then apply that to your specific patient.

So security, don't put this into your generic Google Gemini or your ChatGPT. But that being said, I think the more specific risks are it can get it wrong, and so trust but verify. And I've heard people refer to these models as functioning at the level of, like, a really good intern. So- We hand-- We used to hand the interns the four hundred pages of outside hospital records and say, "Tell me what's important here."

And it's kind of reasonable to do that to one of these large language model tools. When the intern comes to report out to you, you ask questions, right? You confirm that what they're telling you makes sense. And I think the potential risk of large language models is we've been working with interns for a long time.

We kind of know what their [00:06:00] capabilities are, and we have other ways of assessing whether they're giving us accurate information. We haven't been working with these tools for a long time, and many of us don't really understand how to do that error checking, how to check for hallucinations. So, I think it's in evolution, and my cautious answer is yes, I think you should use them.

But as you do start using them, make sure you're checking them as you go, and make sure that you continue to trust what they're providing you.

Trish Kritek: That's so helpful. And I think that part of like, yeah, we would-- you described exactly what I think I did a thousand times as I was, you know, admitting patients.

And the part about I know the questions that I would then ask to kind of trust but verify, is that a skill set that we should teach people? Like, how to then follow up, or is that something that you just learn by trial and error? because honestly, the way I learned it in the people model was trial and error, but I'm curious if there's a way to help people do that.

Jake Doll: I think we need to help people. There will always be super users [00:07:00] who honestly have already integrated these tools throughout their lives, but there are other people who don't know how to even get started. But I think I'm more anxious about people who get started but then maybe don't understand what the risks are or maybe move faster than they should with integrating these tools, and then they expose themselves and their patients to, to risk of misdiagnosis or bad information.

Again, this is a new technology. It's a new tool. As physicians and as providers, we're very used to integrating information sources into our practice. We've been offloading our brain forever, right? Since the Physician Desk Reference after World War II. We've been saying, "There's too much medical knowledge out there.

I can't keep it all in my head. Where's my resource?" But I would argue up until now, we've always had a lot of transparency about what that resource is. If we're referencing a textbook that's fifteen years old, we know that that's a little bit out of date, right? If we

go to an online encyclopedia like UpToDate, like the authors' names are [00:08:00] there and the references are cited.

We've moved to this ability at least to ask a specific question and get a very specific answer back from one of these AI tools without having to then vet that information. We should be, but we don't have to. And so the real risk, I think, is people moving too fast and maybe not knowing how to interpret, understand, and check for accuracy what they're receiving from the tool.

Trish Kritek: Yeah, that seems like really important education, and I know it's an area of focus for you as we move forward with this. I don't think you've said this exact use, but I know that it's something that people are doing a lot more, and it seems important in this one too. So I wonder if you could talk a little bit about, I hear about people trying to use AI tools to help with differential diagnosis, to say, "Okay, here's the story, here's what I know, here's the data.

What should I be thinking is going on?" And, you know, I think there's a lot around that that is [00:09:00] about learning how to do that, what are the risks, and is it really helping us in figuring things out? And so I wondered if you could talk a little bit about that, the use of AI around differential diagnosis or diagnostic strategies.

Jake Doll: Sure. I think this is probably for me the most exciting frontier of these tools, but also the place where probably the most risk lies. I have heard described, and I kind of agree with these descriptions, that at times these models feel like having a trusted specialist on your shoulder. But all specialists, right?

You know the ID doc who you really trust and you would want to curbside for a specific question that maybe doesn't require, like, a full consult. ChatGPT or Google Gemini or OpenEvidence actually serves that role pretty well without needing to call your friend, right? And so I think it's really tempting to start outsourcing more and more of those decisions to the large language models.

What's the data? So there have been a lot of [00:10:00] anecdotes, and there was a case published in the New England Journal using their, their typical format at Mass General where they do the expert diagnostic conference. And they used that same format and published it, but in this case, they included their expert diagnostician, and they gave him two weeks, I think, to look at the data and come up with a diagnosis.

And then they provided the same data to a large language model in real time and asked it to come up with a diagnosis, and within five minutes, it did get to the same diagnosis. It was a different trail of reasoning, which was interesting. Both of them kind of showed their work, but they got to the same place.

And wow, that makes it seem like maybe we should be using this for all patients. However, there was a, there was actually a randomized study where they gave One

group of clinicians, the large language model, to assist with their diagnosis for specific clinical vignettes, and another group just used their normal clinical tools, and there was actually no difference [00:11:00] in their success at reaching a diagnosis by just providing them with an AI tool.

Interestingly, if you took the clinicians out of it and you just asked the AI, it actually did better than both groups. And this was published in JAMA a couple of years ago. But that, I think, raises a lot of questions, but for me, as somebody who's interested in learning and training, one of the key points is here we obviously have a powerful tool, but just handing it to people, at least in that study, didn't fix any problems.

And so if the tool is powerful, it still has to be integrated into our care processes, into our own individual decision-making so that we can actually help patients with it.

Trish Kritek: Since this is a, a podcast, nobody could see that my eyes just bugged out of my head when you said that the AI did better than the clinicians.

And I think it raises this question where I'm going to feel kind of old saying it, but, like, I think people talk about are we going to, like, be out of a profession, or maybe are we just going to take away the stuff that we [00:12:00] actually think are the other really important parts of providing care, which is the people part and the connection part?

And so I know there's maybe not an answer to this, but I'd love for your reflections a little bit on the potential impacts on the profession, which I think there are lots of worries about. And you're making an argument for some benefits. I think there are some clear benefits. I think interns not spending two hours summarizing outside hospital records makes a lot of sense.

But boy, that's a little sobering that maybe the AI is getting to the, quote unquote, "answer" faster than the clinician.

Jake Doll: Well, I'm, I'm so glad you asked that question. It may not be clear from my answers so far, but I'm actually really a techno skeptic. I'm not on social media, and I get really frustrated when tools don't work, and I think it actually makes my life worse when we have bad tech.

And the concerns are real. I'm presenting these as tools, as maybe a supercharged Google search [00:13:00] engine in some cases, but I think the potential place we could end up, which I would call kind of the Star Trek scenario, and that people are really worried about is physicians, providers just end up being an agent for the AI.

So in Star Trek, they go in and they wave like a wand at the patient, and two seconds later it provides a diagnosis and a treatment plan, and they're just there to kind of hold the patient's hand and drive the drama forward on the episode. So, is that likely? You know, probably not in the course of your and my career, but is it possible?

Absolutely. And I think for providers who are working right now- There's a couple concerns. Number one is, am I participating in the demise of my own profession by accepting these tools into our workflow? Is this a slippery slope that is going to end up with us ending in a place where the doctor-patient relationship is eroded, where practice just doesn't look like what I signed up for as an end result?

But then also in the moment, like these tools really should be making our lives better. [00:14:00] They should be reducing the pajama time that we spend on documentation. They should allow us to look at the patient more so that we're not documenting during the visit. But all of us, I think, are, can be rightfully skeptical of prior technology that was also promised to improve our lives and has ended up just adding more cognitive load to our work days.

So the skepticism is appropriate, let me say, and I agree with it. And I do think there's a role for training and education so that people can, for now at least, pick and choose the tools that they think is most helpful for their practice and, and avoid the ones that make them feel icky or doesn't help their patients or doesn't make their lives better.

Trish Kritek: I want to talk more about the education, because obviously that's part of what you do and what you're going to do more of. Before we go there, though, I want to follow up on that thread which you started earlier, which is using ambient AI to support us in our clinical spaces. And, you know, I guess in my heart, I am hoping that it does allow for the [00:15:00] physician or the clinician to sit and, like, actually talk to a patient in a way that's not having to be typing on their computer, and that, as you just said, that there will be less pajama time, that people will be more efficient in making their notes.

And I have two questions about that. I do believe there's some data that supports those things being true, but I'm curious if you know more about that than me. And then secondly, I'm wondering how the patients feel about that part and, and if that's actually effective for the patient. Like, I might feel better as a physician being able to look at my patient because that's what I want to do, and I'm hoping the patient feels better about it too, but I'm not sure about the data on that.

Jake Doll: I can provide a little data on the first one, which is that there have been a couple smaller single center trials and at least one multi-center trial where they've just looked at do clinicians spend less time documenting, and what is their self-reported satisfaction when using an ambient AI documentation [00:16:00] tool.

And in general, I would summarize that there are benefits. It's- Look, we're talking maybe saving minutes to maybe an hour in a typical clinical day, but that could be huge, if you integrate that over the course of a year, much less a lifetime. As far as what patients think, we probably need some really good qualitative data here to understand on a real individual basis how patients interact with this.

I would say that the guidance for how to use these tools always comes with how to appropriately talk to patients about the use of the tools, the need for upfront consent for the tools, and obviously the need to check the output of the tools, especially since all of these reports are available to patients essentially in real time at this point.

And so there is some risk there if, for example, the Ambient AI tool puts out the wrong summary of the clinical encounter, you don't catch it, and then the patient sees it in their, you know, their Epic inbox. [00:17:00] I think that could be a potential challenge. What we can do about that, I think, is, again, going back to the training and education piece, is really being very clear about what the expectations are for the use of these tools.

So yes, consent, yes, prior knowledge for patients, but also the more that we use it effectively as clinicians, I think the more patients will also trust it. It's not going to take too many kind of index bad outcomes though, or, you know, legal cases before people get really nervous about these tools, and so I think starting up front with best practices makes sense rather than responding to problems that happen downstream.

Trish Kritek: Makes sense to me, and I'm curious about, a little bit about-- Let's just talk about the education around this because what I heard you say is, yeah, we should be teaching about consent, and that consent is including the fact that this is using AI, and probably there's best practices in teaching about consenting folks.

And then for sure there's education around you need to [00:18:00] read and edit the results that come out of this, in the same way, to be honest, that if you dictate something, you need to read and edit it. I know that you, because I have many times read people's notes who did not do that, and they say crazy things that sometimes make us chuckle, but also are sometimes problematic because they're not saying what actually you were trying to say.

So those seem like two clear areas of education. Are there other places, let's start with just Ambient AI, that you think, "Yeah, we should really focus on education for folks in this space"?

Jake Doll: If you talk with the technical folks who are in charge of developing and implementing these, they at least express to me two major concerns.

One is, I would summarize as not enough use, and the other is too much use. Or instead of r- use, maybe I should say reliance. So in the not enough reliance category, it is the clinician who uses ambient AI, but then spends half an hour going through and editing it [00:19:00] so that it sounds exactly like their old handwritten notes would've sound like.

And then you've lost any potential efficiency. And in fact, now you're just frustrated because you spend all your time saying, "You know, this AI doesn't know anything." The

flip side of that, obviously, is the one that you highlighted, which is somebody who doesn't check it at all in an extreme version or doesn't check it enough.

So I think some level setting with training is possible there. It is also possible to check the final note versus the original note that the AI put out, and we could even benchmark people to say, "Hey, you're changing 60% of every note, whereas your colleagues are only changing 15% of every note." And I'm just throwing those numbers out.

I don't know actually what the right amount of editing would be. But we can then provide that feedback to people to help them kind of optimize their practice. I think it's also reasonable to have some spot checks for error rates as well, just to make sure that there's not some systematic bias or [00:20:00] hallucination that is present in the AI, and I know the technical partners are working on ways to do that efficiently.

Trish Kritek: That's really cool, actually, and I think those are the kind of data that would help people and the kind of thing that I think we'd love to know. I'd love to know that about my inpatient provider-led charge capture right now as well. So I think that feedback for people is, is really helpful. Let's talk about education about the other uses, which I think are the ones where we're asking it to synthesize or summarize or search key aspects of education and training related to those uses.

Jake Doll: Sure. And some of these uses are really relying on general purpose large language models. And so I do think it's important to step back and just make sure people understand what these tools are, what these large language models are, and what the general risks and benefits of them are. And honestly, people have a lot of concerns about AI that go well beyond, does the tool work for the patient in front of me?

Those concerns [00:21:00] can include, as we've talked about, is it eroding the profession and the doctor-patient relationship, right? Am I heading down a road to a place that I don't ultimately want to be as these tools become more and more prominent in my practice. And then things that are completely outside of medicine, like the use of energy and water for these large data centers, and even the kind of non-zero chance that we are helping create a super intelligence that, you know, at the end of the day is going to be detrimental not just to the profession, but to life on Earth itself.

Like, that is not an unreasonable concern at this point. Just talking about whether ambient AI works is probably not sufficient in this territory. So talking about how these models are developed and trained in general, what they do and do not do well, and then being really honest about these other off-topic things and giving people a space to talk about their own practice.

This is not, at this point, [00:22:00] again, a tool that is going to be turned on or off for everybody at the same time. I'm old enough to have participated in three electronic medical record launches. I'm old enough to have written orders in a paper chart so that dates me a little bit, and there was a day on Monday when we were writing orders in the

chart, and on Tuesday morning that went away and we were writing in an electronic medical record, and you could either do it or you could retire.

And that is not what's happening right now with AI, and so I think giving people all the knowledge they need to be able to select the things they want to do that they feel comfortable with, and then implement them appropriately and safely is where the education and training comes in.

Trish Kritek: That makes a lot of sense, and I appreciate you raising the, the kind of specter of understanding what AI really means, because I think there's a lot of people who, who don't get that and what a large language model is, and, and how it learns and what it's using to learn.

I think all those conversations are helpful as part of education. I appreciate you saying like, and there's bigger picture things like the impact on the planet [00:23:00] that's related to the use of AI. I'm still slightly daunted by the idea of, like, a super intelligence being created, and I heard this piece on NPR the other day where, like, there was a chat room for bots talking to bots and they were planning to take over the world, and that starts freaking me out too much, so I don't want to go there.

But I do appreciate the kind of... I think a lot of us appreciate the potential environmental impact of this that can't be avoided as part of the conversation. So all of that is really helpful. I also wanted to talk about one last thing about education. At least in my experience, I am probably a slow adopter of AI, and I recently used it on a significant project and I found that as I worked with it, I learned a lot, like, about how to ask questions and how to ask for iterations and what to put in to get to the place that I wanted to get to, what worked and what didn't work.

It was mostly trial and error, to be honest. So I'm wondering about that part where we teach people how to do the questions or [00:24:00] the requests or the iterations. Is that something that's teachable?

Jake Doll: It is, and there are programs that will help people think through that. I think it's an area that the tech companies have not done as well as they should.

You will see people say, "Well, just ask the AI what you should be asking AI." And that will work to an extent to help. The AI will actually walk you through the specific questions and processes you should use to use it more efficiency. And that's really cool, right? That's not something we've ever been able to do in humanity before.

But we need more than that, and we need some specific guidance on how to ask those questions. However, I do think that we can't pre-specify all of this, and I'll just give one example from my own practice recently. And it's an example I think of where AI can be really powerful to help us know what we don't know and maybe areas where we can grow both with lifelong [00:25:00] learning and professional development.

And I was recently called in for a cath lab case for a patient who was very sick on ECMO from a PE. And the question for me was whether we should insert a temporary hemodynamic support for the LV with a microaxial pump. If the listeners don't know what any of that stuff is, this is a pretty niche Situation.

I didn't train on this in fellowship. Like, this scenario did not exist. Um, there's not an up-to-date page on this. There's not really a lot of data if you just search Google or PubMed. And I had a general sense just from my experience in using these, you know, devices before and a couple specific cases of what I should do in this scenario to figure out if this was the right way to go.

But as I was sitting there waiting for the case to start and had some extra time, I thought, you know, what does AI think about this? And it confirmed a lot of the things that I was already thinking, and it then provided some pretty specific hemodynamic criteria that I [00:26:00] could consider as thresholds for whether or not to make these decisions.

And then as I asked it, it gave me the actual sources and references that supported its decision-making on that. And for me, I made the final decision, but it gave me more confidence that I was going in the right direction, and it actually sent me down some pathways where I could improve my own knowledge and understanding of these hemodynamic processes.

I think this is really cool and fun in a way that people can start to experiment with AI fairly safely, areas where they maybe don't feel as confident or don't feel like they've updated their knowledge. Ask AI what you're missing, you know? Give it some questions like you might give a trusted colleague and see what comes out, and then iterate on that and ask it for support and advice, and you might be surprised with what you get, and it may be, you know, helpful for you going forward.

Trish Kritek: I think you just inspired me to try that out the next time I'm on clinical service because it's, I think, inspiring to hear it opened up some ideas for you to think about further. It didn't necessarily tell you all [00:27:00] the answers. And I really like the part where it said or maybe you asked, what are the references from which this guidance comes?

And you could actually look at the primary data if you felt like that would be important both to learn but also to maybe validate what the, what it was sharing with you. So it inspired me. Maybe it'll inspire some other folks to just test it out that way. I realize you might have lost all the people when you started talking about ECMO, PE, and new device to offload a ventricle, but the concept hopefully still got through there.

We've talked about a ton. I think you've shared a lot, and I feel like I've learned a lot. I ask everyone at the end of our podcast if they have one last pearl that they want to share with folks. So I'm wondering if there's one more thing that you, you know, I didn't

ask about or you think is really important to emphasize as we think about incorporating AI into our clinical lives.

Jake Doll: One word we haven't said yet, Dr. Kritek, is hallucination, and I think a lot of people are very nervous about AI hallucinations, and that is a reasonable thing to be wary of, and also [00:28:00] is somewhat distinct from just error. So again, one of the completely unique things about this technology is it will answer your question, and, um, it will feel in most cases pretty confident about that.

When you do a Google search, if there's no information out there- It won't return a site for you to look at, right? But AI will figure out an answer for you. So I wanted to address hallucination a little bit. The first thing is that a lot of the hallucination problem that was seen in the earlier models has been ironed out, and by using a tool such as OpenEvidence, which has been trained on scientific data and medical data, it's pretty rare for you to get just a really bad answer.

You should still check, you should follow the links, but you can feel more comfortable in the data than the information than you would have two years ago. In addition, these outputs, at least in my experience, are getting much better at telling you when they don't know. But you can always [00:29:00] ask. Now, it's a little silly to say, "Hey, AI," you know, "give me the answer and don't hallucinate."

Obviously, it's not going to hallucinate. If it knew it was hallucinating, it wouldn't do it. But I think trusting but verifying, following the references, and asking additional questions if the answer does not align with what you already know about the clinical question is really important. But I don't want our listeners who maybe have heard these horror stories about hallucinations to say, "Well, that just taints the entire system."

I do think things are better, and that can be mitigated in a way that makes it safe for our patients and for ourselves.

Trish Kritek: I think that's really helpful information. And I will say, I did tell it, "Don't hallucinate." I don't know if it did anything, but I did say that to it. I think your advice is better than mine, which is, you know, follow the links and trust and verify, but importantly, that the tools are continuing to improve so that risk is not gone but mitigated significantly.

So Jake, thank you so much. [00:30:00] Again, I said this a second ago, but I'm going to say it again. I feel like I'm learning, which I do in every Thrivecast, but I'm learning a lot about AI from you, and I guess I'm going to do a little preview that I think in the future there'll be more and more education about how to use AI.

So I'm hoping people will engage with you and others as we build that out because I think there's so much to learn and so much potentially to gain as we also continue to mitigate the risk. So thank you so much for the thoughtful conversation.

Jake Doll: It's been a real pleasure. Thanks for having me.

Trish Kritek: Of course.

And I'll say as I always do to listeners, if you want to hear more episodes of Thrivecast, you can find them at Apple Podcast, Spotify, or wherever you find your podcasts. You can also find them on the UW School of Medicine faculty website at faculty.uwmedicine.org. Thanks for listening, and have a great day.