AI-GR Pod 42 05.15.26 Travis Zack

[00:00:00] How do these pieces fit together so that you're pulling the right seven references instead of 12 redundant references? So, if you're training a model for that, the actual training set you use is very different than what you would do if all you're trying to do is get the right answer. But it's also a lot more robust because then what you're trying to set up is: what is the right paper at any given point in time, given the day as things change, as opposed to what is the right answer given a question that might change tomorrow, right?

That answer might actually change. So, those were the kind of retrieval questions I was very interested in, but the start of it. And then, of course, you also have to get the answer right, right? So, those are both equally important, but having both of those as pillars, I think, is really critical to thinking about what goes on under the hood in OpenEvidence.

Hi, and welcome to another episode of *NEJM AI Grand Rounds*. I'm Raj Manrai. I'm here with my co-host, Andy Beam, and we have a really exciting and fun conversation to bring you today. This is with Travis Zack, Dr. [00:01:00] Travis Zack, the CMO of OpenEvidence. And so, if you're a physician or clinician practicing, there's a pretty good chance that you've already been using OpenEvidence for some time.

And this was an amazing opportunity for us to just really dig into the background of, not only OpenEvidence, but how Travis thinks about clinical safety, uh, the challenges of sort of growing at this pace and this scale, and to learn about his journey from his sort of early days to joining OpenEvidence and now in a leading role as their CMO.

Andy, I thought Travis was incredibly thoughtful throughout the conversation. I really came back with a sort of new appreciation for not only him, but the work that he's doing at OpenEvidence after this. Yeah, really thoughtful. He was described to me before the episode by someone who I consider to be very smart, to be one of the smartest people that this person knows, and I think he lived up to that billing.

Amazing background, very non-traditional. I think a smart person who has done very, very hard things and been [00:02:00] successful in many different arenas, perhaps none harder than making a very fast-growing health tech company. Three or four years ago, having a company valued at twelve billion dollars that sells directly to clinicians, I would have told you that's impossible.

But Travis has been a big part of that. And hearing how he thinks about both the durability of their moat, the product that they think they're building, which is different than I had thought about it before this conversation, and then what they're building in the future, I thought was super insightful.

And again, very measured takes on everything. Very thoughtful, as you said, Raj. And to your point, probably of high interest to our listeners because, as Travis told us, half of clinicians in the United States currently use OpenEvidence. And so, again, anticipate this being a widely listened-to episode.

The *NEJM AI Grand Rounds* podcast is brought to you by Microsoft, Viz.ai, Lyric, and Elevance Health. We thank them for their support. [00:03:00]

And with that, we bring you our conversation with Travis Zack on *AI Grand Rounds*. Well, Travis, thanks for coming on *AI Grand Rounds*. We're excited to have you today. Yeah, I'm stoked. I can't wait. Cool. Travis, it's really great to have you on the podcast. So, this is a question that we always get started with.

Could you tell us about the training procedure for your own neural network? How did you get interested in artificial intelligence? And what data and experiences led you to where you are today? Sure. I try and not take up the full hour and a half, I suppose. I did my undergrad in condensed matter physics.

I was planning on being a pretty basic science researcher. I went to biophysics for graduate school, interested in single molecule imaging, the crystallography, cryo-EM, scanning tunnel microscopy things. And then actually, during my rotations in graduate school, my mom got breast cancer, and I got interested in doing something a little more practical that might have some real-world impact at some point in the next fifty years.

To that extent, I rotated in some oncology [00:04:00] labs and got interested in algorithm development for genomics. So, worked at the Broad and got the bug for medicine and went back to medical school after that. I was just telling Andy that part of that was exploration. I had been doing statistical modeling work, got interested in machine learning and how that could potentially interact with real world data.

That is, we have the genotype with the genome. I really wanted to understand the phenotype better and how to use machine learning for that. Ran into Andy and Zak and got excited about that. So, decided to do research during medical school residency and fellowship in real-world data and how that pairs with genomics.

So, finished in the HST program there at Harvard Medical School and then did my residency and fellowship here at UCSF. So, that's my long journey, to where I'm at in terms of training. Travis, that's a great recitation of your CV, but let's go back further. Yeah. Like, that's one of the— Okay.

—things we like to do in the opening is how'd you get interested in biophysics as [00:05:00] an undergrad? Yeah. What were you like as a high schooler? Could you take us back a little bit further? Yeah, yeah. Um- What's your origin story? Sure, sure. I was kind of a non-traditional background. I grew up in Lahaina, Hawaii, and, uh, was actually homeschooled for a fair amount of my growing up.

I, you know, moved to California and then to UC Berkeley, and I was very open. I had no idea what I wanted to do. I had a vague idea of what a unit was when I entered undergrad. I was coming in as a philosophy major and, uh, was excited about uh, I still love philosophy. I think a lot of it. You know, it's like big questions.

I like big question stuff. And what actually happened, during undergrad is I couldn't get into the philosophy classes I wanted. I had these breadth requirements. I was in a dorm with a bunch of engineers, and I had to take a physical science. I was like, "I'll just take physics with you guys." And then they laughed at me.

They're like, "Come on, you're a philosophy major. You're not gonna take physics for scientists and engineers." And I was like, "Well, screw you, I'm gonna take this [00:06:00] class, and I'm gonna do great." And it was by far the hardest class I'd ever taken in my entire life. But I don't know, I like hard problems. So, I was like, "You know what?

This is amazing. These are also big questions, kind of fundamentals of how things work." So, ended up going into physics. My parents were not science or engineers or physicians. In undergrad, I actually had no concept that physicians did research. It's just like, you know, did not ... You know, I thought they fixed your knee.

They give you antibiotics for your cold. So, it was definitely not on my radar to do anything. I wanted to do, you know, answer big questions, so I wanted to do research, kind of the academic track. So, got into biophysics. Oh, this is a— okay, I'll tell you one more fun story. I applied to all physics programs, and I had an undergraduate mentor, uh, in physics department who had seen that I had

taken a bunch of biology classes, and I had taken all those biology classes because they had great field trips.

Uh, I was like, you know, I liked invertebrate zoology. I liked the forestry class. We went to go hikes and stuff. [00:07:00] Went on boats and hikes and, um, he's like, "Wow, you're taking a lot of biology. You should apply to one biophysics program." My, I think he said his daughter. I think he, his daughter did biophysics, and, and she really liked it.

And that was John Clarke. He won the Nobel Prize this year. Um, and so, I applied to one, uh, biophysics program, and then I, I toured there. I was like, "Wow, this is amazing. This is exactly what I want to do." And then I had no other op— it was just the one program, so that's where I went. And it, it's the program that when I was in high school, I was touring.

I was in Boston in February, and I remember telling my mom, I'm like, "I will never live here." I was coming from Hawaii, right? I was just like, "I will never live here." The contrast— And then I ended up— between Boston in February and Hawaii, I'm sure, is— Yeah. —stark and so— Yeah, yeah, makes total sense. —yeah, it was like 17 degrees blowing sideways with snow.

Just— And then I lived there for nine years. —it's just tee time, right? Um, but yeah, so. Okay. I, Okay. Well, it was Harvard biophysics first, 'cause I did my PhD first and then went back to medical school. Got it. Cool. Yeah. Um, s- s- super [00:08:00] cool. And, I think maybe, we're gonna dig in, of course, to your work with OpenEvidence and what's been going on there and what, what you see as the next couple of years.

Uh, but maybe you can tell us about how you got involved with OpenEvidence before we sort of dive into the current work. Like, how did you meet the team? I always find these stories fascinating. Like, it's easy to— Yeah. —it's easy to sort of connect the dots backwards, right? But a lot of times people like, you know, you can't really predict exactly which way it's gonna go, and so I, I lo— I love these types— Oh, totally. —of stories.

No, no, it's like— Yeah, yeah. Um, most lucky set of four things in a row to happen in my life. The long story there, I hope you, you got, I don't know if we'll have time to talk any actual things if I—. This is, this is great, Travis. This is great. When I was, when I was transferring, you know, going from end of intern year to residency, um, in medicine, you, you, like, have this,

there are a lot of existential crises in medicine. But there's this, you know, there's this increase in responsibility is multidimensional from intern to resident. And of course, [00:09:00] you're responsible for more patients, you have more ownership, you're running a team. But there's a whole other aspect, which is you are now responsible for the education for your team, right?

You're, you're supposed to come up with how to teach, how to teach your interns, how to teach your students while taking twice as many patients and being responsible for all of them and making sure they don't die. And I've always really loved education and, and it's been part of my identity, so I had been thinking about, "Okay, how am I gonna teach?"

I hated the resident. I thought it was less than ideal for residents to just read the up-to-date page for the patient that they're seeing. You're just like, "I can read the up-to-date page, so I wanna do something different." But I didn't wanna have to do a bunch more studying. So, instead, what I did is I was like, you know what'd be great? Is if I created a retriever model for the CPC cases so that I could basically submit information for the patients on our floor and figure out which CPCs I can teach on, because the pathophysiology in those CPCs are amazing, right?

So, if I can create that [00:10:00] retriever model, that would be awesome. And so, you know, I, and the person, you know, Gurpreet, who was, who was my PCP instructor, right? He was the one who was my, I did continuity clinic at the VA. This is Dr. Gurpreet Dhaliwal? Yeah. Amazing. Correct. Yeah, yeah. He, he had this, we, we'd— Yeah.

—do something similar for education, but his was— Yeah. —this file system he'd have next to his desk. So, you built the digital version of Gurpreet's CPC-retriever? He had, like, printed out, right? Yeah, yeah. So, he had, like, printed out all these, and he's, you know, he's, he's got, his retrieval, his internal retrieval model is just like a vice, right?

So, like, we'd be talking about a case, he's like, "Oh, let me go back to 2004 and pull this case," and then he'd talk about it from his file cabinet, right? So, I was like, "Okay, I'm gonna create that." And so, long story short, this was early days, like, BERT kind of model stuff. So, I was like, "Okay, I'll build a retriever for that.

Then I will have cool cases to show my interns when it's time to teach." So, through that, I got connected to *The New England Journal* [00:11:00] to do that

in a very basic way. Gurpreet and I published a paper on this kind of idea, you know, pre-GPT. And so, New England, I think, yeah, Pete Solovitz— Yeah.

who you guys know— Just saw him yesterday. I bumped into him randomly. —had some graduate students. Yeah. Yeah. Yeah. He's amazing. Um, he had a couple graduate students interested in a project that used the CPC cases, and Jeff Drazen connected them to me 'cause they're like, "Well, we don't have the system to set you up that way, but there's this MD out there who's, like, very interested in that stuff doing it."

And the first graduate student he sent me was Matt McDermott— Wow. —who I know you guys both know. Um, and then- Matt's understudy was Eric. Yeah. So, then I started working with Eric after Matt kind of moved on from Solovitz's lab, uh, on these kind of ideas of diagnostics, how to train, you know, how to train good ML models for these things using the CPC cases.

Again, pre-GPT explosion stuff. And then Eric was part of the first few guys part of OpenEvidence. And so, [00:12:00] I became like, he would just, as he was building the retriever for OpenEvidence, he would ping me and be like, "You're an MD, what do you think about this answer?" And I'd be like, "This is terrible. I'll never use this.

This is total crap." And then he'd be like, "Okay." And he'd come back a week later and he's like, "How about now?" I'd be like, "Yeah, still pretty much total crap." Uh, and, and so then they kinda sucked me in more and more and I, I kinda, you know, got more consulting and just kinda, uh, getting involved in quality and product and stuff.

I was basically like a Slack hanger-on for a long time. But that, that's how that connection happened. Very long story. That's, I'm sure you can edit it out. That, no, that, I, I don't think we're gonna edit that down 'cause that's honestly, uh, it's amazing, uh, just to sort of hear. It's like there's no way you would've predicted that you would've ended up at OpenEvidence. No.

Totally not, yeah. From sort of starting with *NEJM* CPCs, right? Or starting with— Yeah. —starting with what you were doing. I, I. It was, it was a total si—, you know, my, my research is, was kinda like real world data in oncology, predictive models, understanding toxicity, all these things. And then I had this total side project on medical [00:13:00] education for retrieving from the literature.

That was total passion project. Did not think it would have anything to do with my academic career or anything. It was like, wow, I get to work with Gurpreet talking about the CPCs— Yeah. —and AI? Like, I have to do this. And that kind of created this connection with, with what became OpenAI. I, I have to just, uh, remark, and I think Andy knows a lot about the, like, not a, a lot of this already, but I just have to remark at the sort of, like, wild similarities that you and I have in some of our interests and some of our background.

So, I also studied physics as an undergrad. I then went to HST, and then at HST I became obsessed with the CPCs. And so— Yeah. —in the CPCs it was like, you know, just the way the differential diagnosis would be formed, how you would then get the correct answer, and then how we just, just have this amazing 100-year history that is not just today, but looking back and how diagnoses were approached and what diseases were sort of the final answers back in the ’20s and ’30s, how they sort of blend into the background now, right?

Yeah. And into the presentation of case. And so, uh, I, [00:14:00] you know, I got obsessed with the CPCs, and then we worked on building the system, uh, CaBot, named in homage to— Right. —Richard Cabot, which I think— Yeah. —Travis, we’ve talked about a little bit before. Yeah. And, uh, we published a CPC using the, that CaBot-generated the AI diagnosis of, and this was last fall.

And of course, who did they select as the human diagnostician? One of the best— Yeah. —diagnosticians ever, Gurpreet Dhaliwal, who I didn’t know that you actually knew, also that you had worked with. And then Gurpreet and I, we started talking, and then there was a nice back and forth, and I haven’t actually even had time to tweet this out or anything, but there was a nice back and forth and we got to work with Gurpreet and write up a response to our *NEJM* CPC where, uh, someone was asking back on the topic of CPC retrieval, why didn’t CaBot pull this case from, like, 2019 was similar when it was approaching the case that we were solving, that Gurpreet was solving?

And we had this just sort of amazing back and forth, and Gurpreet has talked about this concept of wayfinding or navigating to the correct diagnosis as opposed to just spitting it out. And so, we [00:15:00] invoked that and we, we essentially talked about, like, you know, it got the, it localized the lesion, right? It recommended the test, and we just got to reflect on sort of what is useful AI for a clinician who’s actually using this for a second opinion or even in other ways, versus what are the ways in which we’re currently benchmarking them, or what do we tend to fixate on, which is, like, final diagnostic accuracy.

And the CPC is actually this great data set or trove of reasoning, right? The sort of reasoning chains in the differential diagnosis, and then— Exactly. Yeah. —by the care teams. And so, I, I just have to remark that, like, we have, like, this, uh, wild sort of similarities and in interests that have, uh— Yeah. —that, that's pretty cool, too.

Yeah. But I, I wanna—. That's awesome. Yeah, go ahead. Yeah. Go ahead. Yeah. I, I was just gonna say, yeah, um, again kind of the project is a little half, it was half-baked at the end because I ended up having—. Turns out that being CMO of a, of a rapidly growing startup takes a bit of time— Yeah. —that I— Yeah, just a little bit. —didn't anticipate.

Yeah. But we had a project similar, uh, you know, with different intent with the idea [00:16:00] this, you know, Socratic guide that takes an agent and walks in an adaptive way through the CPCs. But to your point, you know, the, I think the reason that attracts us so much is not the answer.

It's the— Right. —wayfinding that is properly diagrammed by— Right. —the expert that comes in. Either the idea that the answer is not the important part here, right? Those experts, a lot of, I think we actually, when I was working with Matt, we actually proved that sometime in the early 2000s, they started giving enough more information such that people start getting it— Yeah. —right more often.

Yeah. Um, 'cause but it, it's not about whether they get it right or not that matters. It's about the reasoning process— Right. —that gets the answer. Yeah, and I think, like, one thing led to another, and we were initially focused on, like, writing up the differential diagnosis with the reasoning, then using them as, like, a benchmark for different reasoning tasks that are evidenced in the cases, and then eventually just can we create a model to present a slide-based presentation and narrate it through that physicians can watch that you would see [00:17:00] on— Yeah.

—on Zoom if Gurpreet were dialing in and, uh, presenting his, uh, his deck as sort of the inspiration. Um, okay, I think that's a great transition. So, you were working on the CPCs. You— I think you actually started as an assistant professor, right? You're at UCSF, and you're working more and more with the OpenEvidence team, and then now at some point you join as the CMO, right?

Or you join the team, you become the CMO of OpenEvidence. Yeah. And maybe we can just start with a very practical kind of question just to start digging into the work that OpenEvidence is doing with how, like could you

walk us through how a physician actually uses OpenEvidence at the point of care and maybe compare it to tools like, it almost sounds a little bit silly to ask this now, but there are probably some physicians who haven't used the tool as yet.

But I think through your lens— Yeah. —like how physicians are using this, like what's a typical use case, and maybe also how that is different than some of the other tools that they would reach to that are in their quiver, like UpToDate or a Google search or, or something else, or just ChatGPT. Sure. Yeah, [00:18:00] yeah.

Absolutely. And there are a few distinctions there, and I think it comes down to what we started this conversation at, which is what was the problem we were trying to solve, and how do you train a model for that? You could imagine that starting with this, okay, from the outside one, it'd be okay, there's a medical question.

What I wanna do is train a model to give the right medical answer. And that was not the problem that we started trying to solve, right? The problem that we were trying to solve is, given a medical question, what are the right references to answer that question given today's date and given the person asking it?

Um, with that last one being a little nuanced, we can talk a little about that. But that's, that is a slightly different question. But it also dictates how you train the model, right, in a very strict way, right? And so, our goal is two part, but the first one is, given a medical question, how do we identify the right references to answer that question given the extreme complexity [00:19:00] and multi-dimensions that in the medical literature that exists on, right?

So, one is how relevant is that paper to the question being asked? Uh, two is how recent is that, and is it still up to date? Three is kind of the strength of evidence within the paper, and the last is kind of authority. And I wanna be clear, like, there's nuances and there's actually qualitative information in all of those, right?

And then there's a last piece which has been more and more exciting for us, which is think about rather than think about this as a single retrieval problem, think of this as an optimization problem, like a knapsack problem kind of thing, right? Where, well, actually, how do these pieces fit together so that you're pulling the right seven references instead of 12 redundant references or, you know, 12 references where four of them are redundant.

So, if you're training a model for that, the actual training set you use is very different than what you would do if all you're trying to do is get the right answer. But it's also a lot more robust, because then what you're trying to set up is what is the right paper at any given point in time, given the day as things change, as opposed to what is the right answer given a [00:20:00] question that might change tomorrow, right?

That answer might actually change. So those were the kind of retrieval questions I was very interested in, but the start of it. And then of course you also have to get the answer right, right? So, those are both equally important, but having both of those as pillars I think is really critical to thinking about what goes on under the hood in OpenEvidence.

So, how do doctors use it? In a lot of ways for what doctors see, it's a lot like a chat interface, right? You ask a medical question, you get an answer, but that answer is not just in isolation. It's with all the references that help support why we believe that that is the right answer that you can go then double-check.

We're moving to more and more multimedia, right? So, it's not just text answers. You also have the Kaplan-Meier curves and exactly why they say what they say. You have clinical images for dermatology or x-rays to show, okay, this is how it looks in three different real-world examples. We have the flow charts for how you would practice if you're an oncologist in different areas.

[00:21:00] So, that's how it looks for a physician. How physicians are using it is obviously something I'm really excited and look into. What I like to say is based on our data, doctors are often, like most of the time they're using it, it's to double-check things that they think they already know. It's this idea that as you're practicing in outpatient, inpatient, for every patient you see, there are probably four decisions you have to make, and I bet three of those four decisions you're only 95% sure of.

You're probably 100% sure, but you're not quite there, right? And obviously, if you're 95% sure, you're not gonna spend 20 minutes for every single patient getting to 100% sure. Uh, and then there's a question that maybe you're 60% sure or, or you know, very much less sure, or maybe you have no idea. And what, you know, what we're seeing is a lot of times it's just being used as a double check to make sure I didn't miss anything.

Does the chronic kidney disease actually affect the answer? You know, I haven't seen this kind of patient in four months. Have the guidelines changed? So, those kind of double check questions on [00:22:00] things they're pretty

sure of. And then, you know, there are these kinds of, wow, I haven't thought of IgG4 disease since I was in medical school.

Can you remind me what the heck is going on? Or differential diagnosis stuff. But you know, in our data, things like differential diagnosis, which is stuff *The New York Times* likes to talk about with relate to AI, is really only 7% of questions that get asked by— Is, is most of it management? —doctors on our platform.

Yeah. Yeah. It's treatment and management. It's most, I mean, I know we're not slideshows. I'm a big sl— you know, I can show you the data. You don't have to, uh... So, it's mostly treatment and management around a number of different axes. Um, and it's things like next steps. If it's diagnostics, it's not like I have no idea, give me the differential.

It can be, but that often is trainee dependent, right? Where it's like they do that a lot more as intern to residents than they do as end-of-attendings. And, uh, even the diagnostic stuff, it's I'm thinking of these two things. What are some features that might help me go one direction or another? Like wayfinding.

Or what's the next diagnostic test that will help me distinguish it? Yeah. Yeah, exactly. So that, that's kind of how doctors are using it. With regards to how it [00:23:00] differs from other tools out there, there's a big distinction between what we do in UpToDate in kind of orthogonal ways. And so, and I think it's important to, to kind of think about that.

We're precision medical information retrieval. So, you ask a very specific question, we are going to answer that very specific question. So, that's great from an ergonomics time optimization perspective, right? There are downsides to that, and I, I like, I like to talk about it and then recognizing that and then figuring out ways that we can be better at it.

The downside of precision medical information retrieval is you lose a lot of the residuals you get from a browsing experience. So, if I wanna ask a question about coccidioidomycosis, but I don't really think about coccidioidomycosis, I might ask, "Okay, what's treatment for coccidioidomycosis?" And I'll get very specifics on how to treat.

But if I don't know anything about coccidioidomycosis, I go to UpToDate. I'll probably just type in the [00:24:00] term, and then with that I'll have to go through, okay, what are the symptoms I should be thinking about? What are the differentials? What are the diagnostic steps I have to actually accomplish? You

know, there's a lot more residual information that you have to get through before you get to your specific answer, and that information may trigger different neural firings on what to do next because you're actually being exposed to it in ways that precision medical information search, um, doesn't expose you to.

So, I think that there are benefits to both, right? There's a, so, what we're trying to do in the next steps is give you the precise answer, but then learn from where you're at in your practice setting to suggest better follow-up questions that could get you towards information you may not have thought about, the kind of unknown unknowns in your process where you didn't know that you should know this information. And kind of bring that browsing experience back in a way that's still very targeted using machine learning.

And then with relation to other tools, I think it's [00:25:00] pretty obvious, but, you know, with regard to other gen AI tools, a lot of it is about the information we use, right? So, we are very conscientious about our content, and I mean that in two different ways. One, there's the walled garden, which means we don't take things from sources that are not trustworthy.

We don't take things from Facebook, Reddit, the wider Internet. Um, you know, Chad and, and other foundation models are doing better and better at that, so I, I think that's true. There's another thing that gets discussed a little less, but, you know, there's so much of our system of basing medicine on the evidence that requires good faith.

And what I mean by that is publishers like *The New England Journal* have a business model, and they have to keep that business model running, and that business model relies on people paying for their content in some way, shape, or form. And I know that from a legal perspective, that is being contested by a lot of foundation models, right?

They're like, you know, this is public information. We're just gonna use it whatever [00:26:00] way we see fit. And I think the whole system collapses if publishers that are producing this, especially groups like *New England Journal* and societies, none of that is actually recognized. So, you know, the other piece that I'll say about the way we work with content is we pay for them.

You know, we pay for it in a very transparent way where we wanna make sure that this works for both us and the group that's creating that content because that's very important to continuing this in the next three, five, 10 years. So,

Travis, just to dig into the comparison, 'cause I think a lot of clinicians are using OpenEvidence of course, right?

But a lot of them are probably just typing things into ChatGPT-2 or one of the other models and asking it its thoughts. And so, I think you touched on this, but one of the first experiences you'll have when you use these models, and I think this is true even today, uh, although I think it's less true than it was a year ago and even less true than it was two years ago, is that the baseline or the frontier models tend to hallucinate.

And I think this is still a problem or [00:27:00] confabulate, right? Make up references and get the sort of the veneer of the argument correct or even the paper correct. Yeah. But then, you know, invent who the authors are or change, maybe some more important and salient things about the paper itself or the core argument that was being made by that paper.

And I think, I, I imagine that you spend a lot of time and there's a lot of effort on your team at guarding against hallucinations. And if I'm just to think out loud about how I would design this, I think it's not at one step of the process. I think it's probably at every step of the process. There's risk that comes with the sort of inputs that go into the system.

There's risks that come with the outputs. And then there's ways of gating or guarding what is actually surfaced to the user in a way that it's maybe drawn from a set of trusted references or verified or checked. And all of these are, uh, I think potential ways that I know others who are dealing with chatbots that are facing patients or facing other users, uh, have to, uh, have, have [00:28:00] tried to tackle or are trying to tackle right now.

I guess my question is: How has your lens on the sort of importance of the problem of hallucinations or confabulations, how has this changed over the last couple of years as the sort of core base models have gotten better and, uh, and also as you've sort of been in this period of, of very rapid growth as well, where presumably you're learning a lot about the usage and you know a lot about the usage of these systems in detail?

Totally. Yeah, yeah. I think you enunciated pretty well that, you know, inaccurate information that doesn't have a verifiable source can leak in at many different steps. And the first one comes with inadequate retrieval, which is where you get a lot of these hallucinations for references that don't even exist and/or are not supported by anything, you know, just pulled from weights directly, which has its place, right?

That absolutely has its place in a lot of generative AI. So, inaccurate retrieval can lead to hallucinations. [00:29:00] Absolutely. And then, you know, the second step, and I, I think, so that being said, I think we're — foundation models are getting a lot better at the idea of hallucinating whole cloth references, right?

I feel I actually have, even in the last couple weeks, you know, people have said that they're seeing this in, in different foundation models, blah, blah, blah, but it's much less a problem than it was two years ago, right? But that doesn't resolve the problem of inadequate or inaccurate retrieval itself. And I'm still speaking about hallucinations, right?

So, I'm not talking about inaccurate answers. I'm talking about actual hallucinations where if it's done a search, but it doesn't pull what it needs to pull to answer the question. It will still, due to syncophancy, hallucinate. Um, maybe not hallucinate the reference, 'cause again, it's getting they're getting very good at making sure that doesn't happen.

But still tell you what you wanna hear. But they'll hallucinate— Yeah. —what it says. Which is almost more dangerous, right? Because they don't have the information. It could be more dangerous, yeah. Right. And then, [00:30:00] so, and actually, right. So, I am, that was one thing I wanted to bring up. It's a failure mode that is shared with people, too, to be clear.

Yeah, exactly. Totally. Yeah. So, so we're actually moving to more and more nefarious problems with relation to hallucination. The one of just the reference doesn't exist, you know, yeah, lawyers get caught up in that. You know, obviously MDs could as well, but it is something that's more easily verifiable because Google's actually very good at identifying, okay, can I go pull this paper, right?

And then there's the issue of inadequate, inadequate or inaccurate retrieval, which can force the model to hallucinate if it's trying to answer the question with information it doesn't have. So, that's the next. The, the next two are actually very hard, and those are hallucinations relation to information the model thinks it has but doesn't, and we talk a little bit about that.

And then the last one, which is borderline hallucination and more just reasoning failures, but can look like hallucinations, which is when synthesizing from multiple sources, how to make sure that the model actually reasons and answers a question— [00:31:00] Right. —accurately given the information presented.

And there's a gray line there between is that truly a hallucination versus a reasoning error?

But, but to kind of walk through those. So, given a set of sources, and this is another reason that we feel like it's so important to partner with publishers with full text, is, um, making sure that the answers that we get are not just based on the abstract. Um, because what we've found more and more is anything with problems with the reasoning, often is because we haven't been able to show the full, full text for the article.

And I'll give you a very simple example, but you know, if there's a table four that involves a side effect that involves the rate of rash in a RCT in *New England Journal*, and then the question is, "Oh, what's the rate of rash if given this drug?" If it doesn't have the full access to the FDA label, or if it doesn't have the full access to the *New England Journal* that, that showed that, it might say, "Oh, I have this content.

It must be here somewhere. I'll just make it up." Right? So, that's one where it's like, okay, it actually pulls [00:32:00] the right reference, so it's not incorrect reference, inaccurate reference. But you can still end up with confabulation because even with the right reference, the right information is not present because the model doesn't have access to the right piece of that reference.

So, so that's one, and then the last one I mentioned, which I've been very, and actually, you know, Sam on our team has been very excited to work on, which is this idea of how do you take, let's, let's assume perfect recall where you take all the information for a given, you know, rare disease or, or a specific question.

How do you accurately measure and then actually allow the model to reason across that to create a synthesis that is very accurate given that it has to pull information from multiple different sources? And that is the next step. I think that's where we're, we're at right now with regard to the frontier of language models that is still challenges.

Uh, I still think humans do it better. Where if you take the expert who's the endocrinologist who's been doing the guidelines for 20 years, I mean, I, I know it sounds obvious, but they're gonna do a much better job at [00:33:00] synthesizing what the pros and cons of these 12 papers are and how to create a specific set of recommendations than if you ask a model to do it.

And so, we're kind of trying to learn and, and work on that to think about maybe that's moving further afield to reasoning failures, but they can often look

like hallucinations in the right context. So, I add that to that list of the continuum. And as I was mentioning, in terms of nefariousness, those are the ones that often are, are really hard, right?

Where only the expert who really understands their field and the 12 papers will be able to say, "Yeah, but this is a midterm endpoint for the overall survival. So, even though it's statistically significant in isolation, I really wanna take it with caution because it's not kind of what was the primary objective was," or, or, you know, things like that.

Awesome. Shout out Sam Finlayson, shared friend of the show and, uh, currently working at OpenEvidence on some cool stuff. I wanted to talk a little bit about the rocket ship ride that you've been on. I know that you started this as, like, a hanger-on in Slack and as a consultant. Before we started the show, you mentioned that we had a chance [00:34:00] to chat way back.

I would've made the mistake that you guys did not. Uh, me and Zak worked on a lot of, like, predict the diagnosis versus retrieve the information to support the diagnosis, and one of the things that was never clear to me was like, what's the business model? And it strikes me that you and other companies like Abridge have found meaningful business models for health care AI products.

I think of OpenEvidence as DTC, where the C is direct-to-clinician. Mm-hmm. And it, it's clear that clinicians are willing to pay to access OpenEvidence and be able to have that sort of trusted source, um, of, of information as part of their practice. Maybe you could walk—. Can I, can I,— Please, correct that.

Yeah. Can I skip in just real quick there? Yeah. 'Cause, um, maybe I'll talk, maybe I'll, I'll kind of, um. elucidate a little bit, 'cause I do wanna create one correction, right? Yeah. So—. Maybe willingness to pay is what I should've said because they, they, there's a lot of—. Okay. Good. Willingness to pay.

Okay, good. Yeah, willingness to pay. Willingness to pay. Yeah, yeah. Yes. Good. Okay, I'll let you continue. Thank you. Thank you, yeah, yeah. Um, [00:35:00] my wife's a clinician and, and she, she uses it all the time. Maybe you could just talk, like, walk us through what the trajectory of OpenEvidence has been like and kind of like what it has felt like to go from this kinda like small group of folks in the Cambridge area.

I think to date you guys have raised something like over $700 million at a $12 billion valuation, so just like a complete rocket ship. So, just qual— like from a personal perspective, what has that felt like, and like what lessons have you

learned along the way as the company has scaled so quickly? Oh, those are big topics.

Um, I guess I'll, I'll just spend a little time talking about our trajectory as I see it, again, not as the co-founder, but as a hanger-on early on. The most important thing that I saw early on was there was a focus on how do we make sure that this ends up being a business? But more importantly, first of all, is we'll figure that out.

What is the problem we're trying to solve? And that problem was information retrieval in medicine, [00:36:00] and I think, you know, Daniel and his previous companies had a lot of experience in information retrieval, and he was very passionate about this idea of there's needle in haystack problem in medicine, just like I've seen in other places.

How do we solve that problem in a really robust way? And I know it's such a problem that once we solve it, we'll figure out something to do with it, you know? And, um, we had some partnerships early on that were not this direct-to-consumer model. They were fine. It wasn't that there was any problems with those.

What happened was in the end, Daniel was like, "I want to release this for free to clinicians. That's what I wanna do. I wanna see how it grows. I wanna use that as a flywheel to understand what we can do better, um, because they're the ones, yes, we, you know, Travis, you're great, but the ones who really matter are the everyday clinicians, and we need hundreds and thousands of them to tell us what's wrong."

Um, so, I think that was the big turning point where we're like, we're just gonna release this, um, and [00:37:00] we're gonna get feedback, and we're gonna monitor. I think the other thing that was so much fun is we had an, you know, an email, and we were like, please email us with anything you see. And Eric and I and Zack, you know, we'd look at that email every single morning and be like, okay, this person said there was a problem right here.

We're gonna catalog all these and figure out what was going wrong and how do we make it better. And so that kinda release early, release free, and, uh, just allow, be very open to criticism to make things better, I think really allowed for the kind of rapid growth. I mean, that's incredible, and I think, like, the playbook for software product tech startups but had, had been nonexistent, I would say, in like health care IT startups here, like, uh, up until this point.

Normally, you're trying to land a meeting with the right CIO or something like that. Yeah. So, I think credit to you—. And so I wasn't—. Credit to you guys for like—. Yeah, yeah. Yeah, yeah. No, no, it's not. Yeah, so [00:38:00] I wasn't part of any of those, but the way Daniel, the way I've heard it— Mm-hmm. —maybe I, is either directly from Daniel or, or, you know.

They had a couple of those meetings. Mm-hmm. And Daniel was like: "This is the biggest waste of time. These are not the people I want to benefit from this product. I want doctors to benefit from this product, so we should be giving this to doctors." Again, that's secondhand. I, at that point, I was fortunate enough, I didn't have to be any part of those meetings, but I, I hear it.

Like, and it's like you said, I think it's very much from the software, you know, the, the Silicon Valley playbook, which is you have to get it in front of people who you actually wanna use the tool. Cool. Um, so maybe forward-looking, like, what does growth and success look like for OpenEvidence over the next year or two?

Yeah. Um, so we've been exploring a few different things. Just to call out the elephant in the room, right? We are quickly approaching market saturation in terms of users, right? [00:39:00] Um, we have well over half the doctors in the country. Uh, it's like six hundred and fifty or seven hundred thou... I think we just hit seven hundred thousand clinicians in the United States, uh, who've used us in the last month.

But what fraction of the global market is that? There's always a bigger tam. Right. Yeah, yeah. Yeah, yeah. Okay. I get that. So, so that's, that's part of what we're talking about, right? So that's, that's US. You know, this is a product for clinical use, and so credentialing is very important as part of that, right?

So, if you think about just very strictly dimensions, one is you can increase users, or two, you can increase use, right? So those are the, the two ways to grow. So, with it, with regard to increasing users, international market. And so, to do that, we have to be very thoughtful and careful about it, right?

'Cause we were very careful here, you know. It was, uh, you had to have NPI to log in, and we wanted to make sure that that it was designed for clinicians. We wanted to keep that. So, you know, we're, we started Canada. We're working to, you know, make sure the credentialing works there, and similarly in EU and, and others, one step at a time, [00:40:00] making sure we go through all the security, the privacy, the regulations, so that we can fully expand in those areas.

Cool. So, that's one side, which is, and, and you know, we're, we're doing a lot of fun. Actually a, a project of mine, which, I don't mind shouting out a little bit if you don't mind, is how does this look in different resource settings, right? So, uh, we've been working with Results to Save Lives in Rwanda to think about what is the right way to adapt OpenEvidence when there are resource constraints, like in low-, middle-income countries.

And that's a really challenging topic with a lot of ethical quandaries because on the one side, the right answer is potentially the right answer anywhere, right? You know, evidence-based literature, the goal is we're all humans. They all, you know, we do a randomized control trial, and for all its faults, the idea is this is expandable.

But at the same time, if we're constantly answering oncology questions with $200,000-a-year drugs that are not available in that resource constraint, that's not very useful to the clinician, right? So, how do you adapt for low-, middle-income countries with that balance in mind? [00:41:00] So, that's another thing we're doing in that regard.

So, that's one. And then the other one that's been really fun for me to think about is what are some other ways that we can build for use? So, that is, how do we make OpenEvidence more useful? So, more users, more useful. Um, and so, we're building a lot of things both within the health system, so we're now full circle, um, now working with health systems.

Be like, "Okay, it's great for your doctors. Your doctors love using us. What are some ways that we can help the health system?" And if that involves, you know, AI scribe capabilities with evidence, um, built in, if that involves kind of more inpatient management and understanding evidence gaps in those.

If it involves, you know, a simple thing that we've done that has actually been really effective is taking information from the health system about how they wanna practice, right? Like, um, antibiotic sensitivities, classic example, right? Like, yes, the IDSA says this is what you do for CAP, but the infectious disease department at a hospital knows what the drug sensitivities are for their local strains.

We should make sure that that's [00:42:00] incorporated so when a doctor asks that question, we can provide that. So, those are kind of some of the early things that we're doing in the next, one to two years is always way too long a timeline. Totally. But let's say the next four to six months. Yeah, the next, um, one to two hours, um, is probably— Yeah, exactly. —the right timeline.

One last question before we move to lightning round. I'm obligated to ask this to all startups. I ask it of myself all the time. What is your moat? I think that the frontier, the space, the surface area covered by the frontier models grows all the time. A lot of them are big in health care and life sciences.

Yep. How do you think about using the wedge that you've created to create, like, a durable moat and a sustainable product in light of the sort of exogenous forces that I'm sure you think about all the time? Yep, totally. And so, I think the two modes that we've created, one is based on effort and the other is based on we started early, basically early effects.

And one is, you know, the moat that I talked about with working directly with the people creating the content so that we're using it responsibly. And what I mean by that is, we haven't [00:43:00] quite crossed the Rubicon yet where, yes, I know foundation models are stealing anything. They probably are watching my screen right now, right?

They're stealing everything available. But in terms of reproducing it, they haven't quite crossed that Rubicon of reproducing figures, reproducing images, reproducing surgical videos that someone's created and edited. And so, by licensing that content and being able to provide that, we are visual beings, right?

And it's one thing to see the full statistics of the whole randomized controlled clinical trial, but it's much easier for me to understand where the gap starts to develop if I get to see the Kaplan-Meier curve. Or if I'm an ED doc and I'm looking at a hairline fracture, it's really nice for me to be able to see, here's exactly what that image looks like across a number of distributions.

And so, one is because we started working responsibly with these groups, we've heard from a lot of them that they are very reticent in the idea of working with these foundation models because of, you know, they're starting from a place of begging forgiveness. So that's, that's one. And then [00:44:00] two, of course, is the data flywheel, right?

We have 200 million questions. Obviously, foundation models have a lot of questions, too. The difference is those questions are often asked by consumers, and the questions asked to our model are very, very complex questions, and they're tied directly to the specialist, where we know this is how long they've been training, this is where they practice, this is their specialty, they're a subspecialist in GI oncology, et cetera.

And, you know, we never share that information outside our system, but it does help us in a lot of ways learn both for retrieval, but also for writing on how to make our product better. So, those would be the two that I'd point to immediately. Travis, can I just ask, it's kind of related to what you just answered, but I think it, it ties to maybe some of our, our conversation from earlier.

I think what's amazing is that I walk around the Harvard hospitals and I go to the residents' room or go see my colleagues there, and a lot of them have OpenEvidence open on their phones or on the computer screens there, or Chat or [00:45:00] another, uh, LLM. And it's not like they had like a class that taught them how to use it, right?

I think the medical schools and the fellowship and residency programs, they're like very slowly contemplating with committees how to integrate this stuff, and then committees on committees how to integrate this stuff into care. And the doctors aren't waiting, right? The clinicians who are actually seeing patients are not waiting, and then the patients are not waiting either.

They're, uh, I don't think they're using OpenEvidence unless they're a doctor and a patient, right? But they're using Chat or one of the other models. So, sort of a two-part question, and they're, they're kind of related to each other, but, you know, maybe not. The, the first is, how do you imagine medical school changing?

How should it change in this AI era? Uh, should we have classes on how to use LLMs and OpenEvidence, or are they gonna figure that out on their own? What should physicians learn about AI tools while they're in medical school to prepare them for their future clinical life? If anything. I mean, I, I think a reasonable answer is like, yeah. Yeah.

Well, uh, so I, I think there's a more [00:46:00] fundamental question about what, what should we, what should we be learning in medicine. And then, you know, as part of that, you know, how, how to—. University. How does the fact that I'll be practicing with an AI tool, um, fit into that? And we have the same background, so I, I think you won't be surprised that I'm gonna lean on that bias to say— Physics. —that I think we should be—. Everybody should be learning physics.

Yeah, we should be learning more, more basic pathophysiology, I'll say. I will, I'll go a couple more levels of abstraction than that. But I think the basic pathophysiology is the stuff that will never change. Um, and I think it's, it's

really fundamental to understanding how everything else you're gonna be doing in medicine works.

And so, I, you know, so maybe not the Krebs cycle, you know, but, you know, things like why does an ACE inhibitor work, right? Uh, for example. Or like, you know, the RAS system more generally, right? 'Cause in the end, I don't think it's probably gonna be very useful to memorize second line hypertensives in a CKD patient because, one, we can look that up, and two, the shelf life for that information is just so short, right?

It's, it doesn't, it [00:47:00] doesn't last very long. But, um, I do think it's important to understand, you know, the potential reasons why these side effects happen or the potential reasons why this treatment might work because those are foundational concepts that help you interact and ask questions and, and get to the right answer and know the answer's right, understand that right answer, know when maybe you didn't provide enough context to an LLM to get that kind of specific treatment management one.

Mm-hmm. So, that's from a global perspective, you know, what, what I think we should be teaching more in medical school. And so, that leads to my next thought on what we should be teaching about AI. And, uh, again, biases disclosure, Sam and I created an open source education— Nice. —tool for this. But the idea of that lecture series was, look, if you're an orthopedic surgeon, you may never need to know about, you know, the kidney that much.

But in medical school, you still have to take a couple lectures on the glomerulus. No, no offense to orthopedic surgeons. Maybe, maybe they have to use it a lot. But, you [00:48:00] know, there are these foundational pathophysio- or physiology and pathophysiology concepts that you need to know because we have a very interconnected system, and it turns out that your patient may be on one of these drugs, and you're gonna stress them during surgery, and this is something that may come up, right?

You never know. And I think very similarly, understanding the physiology and understanding the pathophysiology of AI systems is very important. So, no, uh, you don't need to understand the different pathways of a transformer model. But it's good to understand what pre-training is. It's good to understand where that data comes from in pre-training, right?

It's really good to understand what post-training is, you know, in a, in a very basic sense. You know, like what is reinforcement learning? Why, what is the human feedbacking portion of that? And understanding that physiology helps

you understand the pathophysiology of ways this can break and what to be on the lookout for and what are the, the potential risks of adapting out of distribution, for example, or, or things that you might be doing, um, and not [00:49:00] even be aware of it if you don't have that physiology and pathophysiology background.

So, and, and maybe it's prompt engineering. You know, someone else might say, "You know what? All we need is prompt engineering. Just make sure they know how to interact with it, and they'll be fine." But for me, I think one of the great things about medical school is it isn't a technical college in, in some ways, right?

Like, it makes you, you first have to understand that we still have this belief you have to understand the basic science. You have to go to O chem at some point. You have to learn something about the basic physiology. And I think now that we're gonna be interacting with these AI systems, either in an active mode, um, where you're aware of it or more and more probably in passive modes, right, where it's an operation support, and you don't even know it's happening underneath the screen.

I think that, kind of, basic physiology and pathophysiology of AI is really important. It, it's a great answer. I think it's a little bit different than what I, I would've guessed, but I—, Oh! —I, I think it's a great answer, and I think it, uh, uh, to me, uh, yeah. If I, if I didn't answer the question, I'm happy [00:50:00] to, to you to ask again.

No, you total, you totally answered the question. Okay. I, I essentially, was asking how medical school should evolve in the AI era. And I think you, you know, if I'm paraphrasing, I think you said there's a lot that medical school actually is doing correctly already, which is giving you the sort of understanding of pathophysiology and sort of mental schema, a way of organizing knowledge and knowing sort of what threads to pull on for, uh, for accessing information, uh, relevant to your patient.

And of course, the Krebs cycle got mentioned, which is great. But I think people are imagining very different, like very different sort of futures from a very practical training on prompt engineering like you're saying or using the tools to, uh, I think trying to turn physicians into computer scientists in the middle of medical school, which is maybe the, the other extreme.

And so, I, I think, yeah. The, the other, the other thing I'll mention, sorry, the other thing that why I feel strongly about this is, like, another phenomenal thing

about medical school is it actually allows a lot of [00:51:00] maintenance of dedifferentiation, right? Pluripotency? Or like... Like pluripotency. Yeah. Yeah.

Yeah. Right? So, like, if we are like, "Okay, we're gonna make medical school now just how to interact with prompts so that you can use OpenEvidence to treat your patient," we're never gonna have the physician scientists making new drugs. Yeah. Or we're never gonna have the physician scientists who are running clinical trials to test those drugs, right?

Or we're never gonna have the people who understand the business of a health system and how that works, you know? And with boots on the ground. So anyway, those are my thoughts. You know, I would love to maintain that pluripotency of medical school because those foundations allow for doctors who do many things, not just one thing.

Amazing, Travis. I, I think, um, I think that brings us to the lightning round, Andy, right? Yeah, awesome. So, we're gonna hop to the lightning round now.

These are gonna be a bunch of rapid-fire questions. Oh, [00:52:00] boy. Keep, keep your answers brief, and we think if they, if we think they require further elaboration, we'll ask you for 'em. So, one that we like to start with, just because it tends to be, like, more revealing than you think it might be, is, um, what was your first job?

Well, I worked in Lahaina. I grew up in Lahaina. My first job, I got paid to be a magician assistant, um, where I juggled and rode a unicycle for a July 4th parade. That's pretty cool. See, that's why, that's why this is a great question. Like, that's why this is a great question. Can, can you still juggle?

Is it like juggling, like yeah, is it like— Yeah, I mean—. once you can juggle you can juggle, right? Juggling, yeah, I probably couldn't, uh, idle on a unicycle. Okay. Probably could ride a unicycle, but yeah, it's been a long time. The thing is, the, the taller you get, the worse balance is, you know? It's just like there's a biomechanical property there where surfing and unicycle riding just aren't as easy anymore.

Amazing. Alright, here's our next one. If you could instantly digitize one thing in medicine that's still done on paper, what would it be? [00:53:00] Probably pharmacy, uh, prescriptions. Um, I mean, that... I, I'm just trying to think of what I use paper for anymore. Yeah, not that much. Yeah, taking, taking, taking your prescription to Walgreens still seems really stupid.

I know you don't always have to do that, but seems like it's pretty frequent. Okay, cool. Cool. Yep. Um, so this is a callback, and, uh, you can choose, um, to answer one part of the question or other part of the question depending on how much you remember of your first year of undergrad. Who is your favorite philosopher, or what is your favorite topic in philosophy?

Oh, yeah. I mean, uh, Dostoevsky or, depending on how broad you wanna get, uh, Herman Melville. Okay. Oh, wow, great choice. Dark horse there for favorite philosopher. Yeah. Awesome. Travis, what is your favorite non-medical use of AI? Um, I've had a lot of fun, um, giving people different facial hair for, um, lab members.

Nice. Awesome. [00:54:00] Um, this is a question addressed to your role as CMO of a very successful startup. What is your least favorite part of fundraising? Um, sorry. I, I, you know, I'm gonna have to pass on that, but I'll tell you a very good reason, 'cause I don't have to do any of it. Oh, okay. So that's my favorite part.

So, okay. Can I say my favorite part? Yeah. Yeah, that's your favorite part. That's pretty cool. Yeah, fair enough. Pretty cool. That's, yeah, fair enough. Yeah. I mean, you know, we have Daniel and Zack. Yeah and it's just incredible. I just saw— They just let me focus. —I just saw Andy's facial expression at the same time.

Lucky, lucky you. Lucky you. Yeah. Amazing. No, I mean, I, I, they let me focus. They probably don't want me anywhere near that. Fair enough. I focus on clinical, like, how to make the product better, how to make the partnerships better, how to make the quality better. Andy, what's your least favorite part of fundraising?

I also would like to pass on the question. Alright, bye. Thank you. Alright. We'll go to the next one. Travis, uh, this is the next one. Actually, this is the last one in the lightning round. Who's your scientific hero, or whose lab would you join today if you were starting grad school? Oh, [00:55:00] wow. Oh, man. That's really tough.

Um, well, uh, I'll answer two separate ones. You know, um, my sci- I, I had great mentors, so, you know, my mentor in graduate school was Rameen Beroukhim. He's phenomenal, just smartest, nicest person I ever met, and then Eli Van Allen, which I can say the same. Um, and then, you know, I got to work with Atul, uh, for eight years here.

Um, and the amount of lives he impacted, um, on a personal level, uh, was just incredible to me. You know, um, he, oh, sorry. It's a lightning round. Anyway, so Atul is, uh, you know, the, the, the mentor kind of, you know, the person I look up to the most in science. Um, he passed about a year ago, actually, um, very close to a year ago, so obviously wouldn't join his lab.

Um, but I think I'd probably I'd love to go back to kind of one of the, the kind of computer science labs and still learn a bit. Uh, I never was formally computer [00:56:00] science trained, um, so, you know, maybe even Pete Solovitz lab or someone like that. I don't know if he still takes graduate students, but, um. Amazing.

Amazing. Thank you, Travis. I think that's the, uh, end of the lightning round, so that was, uh, absolutely fantastic. I'm very, very bad at lightning rounds. You did amazing. You can tell. Like, I probably. Yeah. Uh, every. My, my shortest answer is probably like four minutes. Yeah, no, you did great. Yeah. Um, cool. So, I think, uh, the last, uh, little bit here, we like to zoom out and ask some bigger picture questions.

You mentioned correctly that making predictions is hard, especially about the future, and doubly so in AI. Um, I think we're like four years post GPT-4 or GPT-3, I forget, but we're still like very early, as they say. What do you think the next four years are gonna look like in AI for medicine?

Do you think, actually, and let me give you a foothold on that, that big question. Do you think change is gonna come fast in health care, or do you think it's gonna be slower? Yeah, I mean, it's already been faster [00:57:00] than most other fields, right? Like, it's been us and it's been the two fields I work in, right, which is software engineering and, uh, medicine, I feel like have had the biggest changes in how people practice. That's interesting.

Or how people do their jobs. Software engineering, totally agree. But if I look at non-clinical things like drug discovery, like, like— Hmm. Yeah. —AlphaFold, like, things like that feel radically different, whereas like health care feels like it's starting to change, but maybe not changing as fast as law, where there's like lower barriers to entry and people can use an OpenEvidence like Harvey for law or, I, it just feels like there's some viscosity in the health care system.

Maybe. There always has been, but like with OpenEvidence, you know, we now have, you know, like I said, over 50% of doctors using us. Yeah. Completely DTC, as you mentioned, right? Yeah. Like, that kind of pace in the last 18

months or two years is pretty incredible to reach saturation with an AI tool that is being used at the point of care to make clinical decisions.

So. That, that's pretty fast. Agreed on [00:58:00] adoption. But have things fundamentally changed as a result? Great question. So, then any, but, you know, there are other, uh, OpenEvidence is one, and I would say that there's been a lot of change in practice thinking about primary literature with that. AI scribes, another one, right, where, um, there's been rapid adoption across, you know, what is now like part of the onboard or part of the offer letter is like, "Do I get to use an AI scribe for, for what I do clinically?"

Right? Because it's so important for, for what, how people practice. So, I think in the next year we're gonna see a lot more around operations, so things that are going on behind the scenes for good or ill. Like I, I have a lot of concerns about AI tools for operation management.

But, uh, hospitals are under the gun in a lot of different ways, and they'll probably adopt those things. Maybe I'll ask, flip it back to you. So, you actually think that in science generally there's been faster adoption than in health care? Um, I think in certain areas of drug discovery, generative models have fundamentally changed the way [00:59:00] that drugs are designed.

Like, you still have to get to the clinic, but the way that you go after, like, a new target, that timeline has been compressed from, like, five to seven years down to, like, less than a year. Sure. In many cases. Yeah, totally. And the way that the work is done, it feels qualitatively different than the way that it, it had been done.

That's a small slice of the drug development pipeline. No, no. I mean, uh, I, I'm, I'm, uh, I'm very excited that you're so excited about it. I, likewise. Like, your answer, I'm like, "Oh, that's awesome." 'Cause I thought you were gonna say. Right. Yeah, exactly. 'Cause you have much, you have much, uh, you're much, much closer to the target than I am, for sure, in that realm.

My other thought with drug discovery is I, I just feel like there have been rumblings on this for a longer stretch, right? People, you know, like you mentioned AlphaFold, right? This is pre, uh, generative language models. Um, people have been trying to do this, and I think it's just, you know, now we have the tools to actually do it very well, whereas pre-GPT, nobody was talking about this kind of stuff outside of Dragon.

Um, so, [01:00:00] so from 0 to 60 feels like a little faster. But I agree. You know, I mean, it's hitting a lot of different industries. There, outside of, uh, general contractors, most people are worried about what my job looks like— Mm-hmm. —in the era of AI. Yeah. Uh, it's a great point, and I think, uh, I, I totally see where you're, you're coming from.

It's like the, maybe the absolute velocity isn't as high, but the derivative, like the acceleration probably in health care— Yeah. —is surprisingly high. Exactly. Exactly. Yeah. Totally. Yeah. Travis, I think the, as, as you've pointed out and, you know, as we've just seen by just talking to my colleagues and, uh, talking to the residents around here, uh, there really is very fast, uh, and very sort of durable adoption by clinicians of these tools in, in clinical practice.

But what I think I haven't seen yet is sort of a RCT or, and I, I know that this would be actually very hard to conduct and, uh, difficult to sort of get off the ground, but an RCT or a real sort of clinical appraisal against a hard outcome of how these tools impact clinical practice. You know, from your perspective, do you [01:01:00] think and I think this kind of echoes one of Andy's questions just a moment ago, right?

There's really, there's a lot of adoption, but, like, how practice is changing. Are errors going down? Are better decisions being made? Is management that saves lives happening earlier? Maybe as a, this is a little bit of an unfair final question, but can you maybe just give us sort of your thoughts about what we need for this kind of next phase beyond adoption by physicians of the products to really have the sort of clinical-grade evidence of the impact of these tools on care?

And the scribe analogy actually is pretty apt because there's also been fast adoption there, and now we're starting to see some of the RCTs. And I think we're learning that, about the, you know, time savings, maybe the lack of time savings from some of these tools, uh, and also the impact on physician well-being and, and other, uh, other measures that, uh, that, that mean a lot.

So, uh, uh, do we need similar types of studies for, uh, for OpenEvidence and for, for tools like that? Yeah. Let me, let me hit a couple of those points 'cause, you know, I'm not a, I'm [01:02:00] not an AI scribe person. We have an AI scribe, but, you know, it's not, um, my focus. But yeah, I was actually just on a panel with Karan Singhal, um, at OAI, and he said something I really love.

We were talking about graduate school and what graduate student, you know, projects focus on, or at least that's how he took it. And I loved what he said. He was like, "You know, I, these kind of AI in health care projects for graduate

students take two flavors. One is betting against the AI and one is betting on the AI, right?"

Like, and I, it's a bit simplistic, but the idea is like, is your big finding that AI sucks or is your big finding AI is transformative, right? And he's like, "I would bet on the graduate student projects that bet on AI," which I think is very interesting because AI is only gonna get better. So, for example, in the AI scribe solutions, yeah, I've also been underwhelmed by how much time they save and the errors and, you know, the, the corrections that have to be made and things like that.

But [01:03:00] physicians adopt it and they like to use it, and it's only gonna get better, right? These, these things are only gonna improve. And my guess is if we did this in a year, those RCTs that show there's not much improvement are gonna be outdated by scribe options that show, okay, there are gonna be. So that's what I'll, you know, I would say that I'm betting on AI.

I think that it's only gonna get better, and if physicians are enjoying, you know, using it, here's where I'll transition. These adoptions, with some exceptions, have not been driven from the top down, right? This is not like you have to use this AI scribe tool 'cause it's good for us. Physicians are choosing to use these tools.

And so, if you're choosing to use this tool for your job, I kind of see it's useful in two ways. One, it makes your job easier, or two, it makes you better at your job. And in my mind, AI scribes make your job easier and make your life easier. Uh, at least physicians have felt that they make their life easier or they would stop using them, right?

Uh, more or less. I, I'm speaking broad generalities, right? [01:04:00] And I actually think that physicians use OpenEvidence, and we have less data to show that they make their lives easier if you do it in isolation. I think they make them better at their jobs because they are looking up and finding evidence where in absence of OpenEvidence, they would just choose to go with their gut.

So, they basically pull from weights versus actually retrieve, their weights being their internal weights. Um, and so, you know, yes, we're still working on, okay, how do we make that pairing, right? 'Cause, uh, a lot of, we're, we're EHR external, right? So, when the doctor asks us a question, we don't know who the patient is that they're looking at, more or less.

You know, we do, we're starting to create partnerships with health systems. But in terms of looking at outcomes, looking at changes of behavior, we're a little bit on the outside looking in, and we're very excited about exploring those. But my thing is, you know, nobody's forcing doctors to use OpenEvidence.

So, in large part, they think they're, you know, they think that this is either making their job, life easier or [01:05:00] making them better at their jobs. And so, agree with you. Um, we'll see in the next year, you know, what we can do in terms of health system partnerships to actually prove that out with primary outcomes.

Awesome. Well, I think that's a great note to end on. Travis, thanks so much for coming on *AI Grand Rounds*. This was really fantastic. Yep. Thanks for having me. Yeah, really awesome. Thanks for, like, lots of really thoughtful answers and, um, loved, uh, getting the opportunity to chat. Yeah. Cool.