AI-GR Podcast 26 01.14.25 Zak Kohane

[00:00:00] I went and I created 1,000 synthetic patient histories, short ones. I then went through, picked 200 random ones, and myself decided, as a human medical expert, who should be seen today, and who will have to be waiting a week or two weeks to be seen. And I did it, and I also weighted which were the easy ones, and which were the hard ones to decide.

And then, I asked GPT-4.0, Gemini Advanced, and Claude Sonnet 3.5 to decide themselves. I measured two things, how concordant they were with me, and how consistent they were with themselves. Because, you know, when you have a doctor, and he's, or she, is accurate on average, but is wildly different between different visits, you're gonna be a little bit worried.

I then said, how can I move these models? I gave them a great sample of other pairs that it would not see, and said, this is how I, the expert, chose. Please use this as a template. [00:01:00] And some, like GPT-4, became more concordant with me, and more focused, more consistent. Others, like Claude, actually did got worse, and actually more scattered.

Welcome to another episode of *NEJM AI Grand Rounds*. This was a really special one. This is our first episode with a repeat guest, and it is our Editor-in-Chief, Zak Kohane. Andy, I don't know where to start with this one. This was a really, really fun conversation that really touched on many, many different things.

I think it's fair to say. Yeah, I think that's right, Raj. This feels just like a fun snapshot into what life as a postdoc was like for us. Freewheeling conversation. Maybe less scotch during our postdoc, but I don't know if even that's right. I think too, I had the added disadvantage of being extremely jet lagged during this one.

So, Zak really pinned me to the [00:02:00] wall. That's a lot of, that's a lot of, a lot of caveats, man. That's a lot of caveats, but it's always fun. We looked back on the year that was at *NEJM AI,* lots of great papers. And again, it's always fun to sit down with Zak and do these end-of-the-year conversations.

I totally agree. I think you said it right. This was a good window into our postdoc time. And honestly, it was a good window into Zak's mentoring. I think there were a couple times where he really sort of recalibrated us and, uh, you know, kind of put us in our place a little bit. And I think it was a lot of fun and a really, just great conversation overall about AI and medicine, and life with Zak.

So, it was fantastic. And I think this is going to be one of those things that I look forward to every year. It's just such a fun thing to do. Hopefully, listeners find it fun, too. Totally agree. 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 Zak Kohane. Alright, Zak Kohane, thanks for coming back to *AI Grand Rounds*. I think you're our first repeat guest. Well, yes, you're sort of obliged to, aren't you? And this is in our contracts. We had to have you on every year. It's true. And sadly, I still think you're the most famous part of *NEJM AI*.

This cannot stand. But I'll keep on coming as long as necessary to make sure that the right thing happens. Alright, 2025, we'll dethrone *NEJM AI Grand Rounds*. That is right. So, Andy is right, Zak. This is— Although you guys are going to have a lot to recover from, from how you were schooled by Larry Summers.

Oh, he dunked on me several times. If you're going to get, you're going to get dunked on by someone, you want it to be Larry Summers. It's an honor, actually. It's an honor. Yeah. But you were dunked on. I admit that. Yes. So, Zak, this [00:04:00] is Andy's right. This is the first episode of the repeat guest. So, this is a question that is inspired by our usual opener, but we are going to ask to you for the first time as our first repeat guest: what major weight updates have happened to Zak Kohane's neural network since you were last on *AI Grand Rounds* one year ago?

The major weight update. It's pretty clear, actually. It's that I no longer feel like we, I mean, the three of us are saying medicine is in deep trouble. I'm hearing it now everywhere. And not just from our large health care systems who have large revenues that are slowly increasing, but employment costs that are skyrocketing.

I'm not only hearing it from dissatisfied young trainees. I'm hearing it from leaders at the [00:05:00] National Health Service in the U.K. I'm hearing it in the, in France. I'm hearing it in the United States. And the question is not, will AI replace doctors? It's, please, can AI help bridge the gap? We don't have enough frontline doctors.

Yeah, I mean I feel this acutely just like having been in the sidecar to medicine for the last 10 years. And so, when Kristen was in med school, I was like AI is gonna come. And now there's like really like, please, like, we need this faster. Like, we don't have enough hands to be able to do our jobs. I do think you're right and that has really happened. It feels like over the last year. And I don't know if the fault lines have gotten bigger, or just there's a bigger willingness to recognize that.

Like, I don't have it, or if it's just like, yeah. So, so you know, in medicine, we like to talk about how a lot of our organs have huge reserve capacity. [00:06:00] And so when you have your organ being attacked, you don't notice it until it cuts into your reserved capacity. So, forgive my teachers for me getting this wrong, but I think something like you survive on 20% of your lung capacity.

And you're breathing okay until then. But then when you hit there, you're short of oxygen. And you start suffering. And I think that's just happened. So, the fact that you can't get a primary care doctor to see you at Mass General Brigham makes the disease now manifest. We're cut to the bone. I think that's a good actual transition point.

So, the first thing we want to start with is actually a perspective that you published pretty recently, maybe a couple months ago, that was in the *New England Journal of Medicine*. And it's compared with what, and then, I really think it is a very powerful essay about what the existing state of the art is in medicine, and where we have access, [00:07:00] where we're completely lacking access, what the problems are with our existing system, as being the relevant thing to compare AI to and to compare new models, new algorithms with.

And so maybe we can start with that one. Can you summarize, what motivated you to write that? And then I'll tell you about some reactions I've heard from folks both here and around the world. Well, in some sense, I was prepared for this feeling by an old phrase from Seymour Papert, one of Minsky's colleagues, the LEGO professor at MIT, who coined this term, the superhuman human paradox.

And this was probably back in the 70s or 80s. And the point he made was that, why are we comparing AI to a super expert where most of us operate in average mode? So just making everybody as good as average, [00:08:00] would be making 50% of the population, if it's normally distributed, better. I appreciate the clarification between the mean and median for not normally distributed.

Since I have a stat nerd here, I refuse to be corrected by a stat nerd. I was about to well actually you. So, I was prepped for that, and then when it became evident to me, and I think as you get older, unfortunately, more and more of your friends end up in hospitals. And therefore, you get more and more stories from them how we are falling short.

At the same time, you have your colleagues who are in medicine telling you that they do not have time to think. That they do not have time to discuss patients. And you see that confluence resulting in a lot of bad stories. So, thinking in my role as a member of the editorial board here at *NEJM AI*, I said, [00:09:00] what should we really be asking in terms of helping the public?

Do we want to say this is better, better than the best doctor you could find? Or is it as good as a really good doctor, and therefore maybe better than half the doctors? And that was the insight because on the one hand, there is a lot of appropriate concern.

Errors, incompleteness, lack of common sense knowledge, hallucinations. But guess what? Doctors, human beings, do those, too. So, when the question is not an academic one, who performs best, but a social one, how do we deliver the best care? It seems more and more plausible that in the absence of a primary care doctor who is well slept, fully caffeinated, and not stressed, maybe we should have a PA or a nurse practitioner, physician assistant or nurse practitioner assisted by [00:10:00] the AI.

They can use, the human can use their common sense, their training, their EQ, and they can be complemented by the EQ of the, of their AI assistant. Remember all things I forgot, weird zebras that may or may not be relevant to the diagnosis. I think that becomes the question. And if we were always going to ask, can this beat all doctors, which maybe one day AI will, we would not be serving our society well today.

Is part of the lack of uptake of that kind of thing at all related to like licensure and scope of practice for other health care workers or are there structural reasons why we can't? Of course, there's so many structural reasons and you've heard me rant about this before. You know, what if your AI tells you not to do that MRI.

That's a few thousand dollars that the hospital's not going to be able to bill for. What if it's going to suggest a [00:11:00] therapy that is more effective but less costly? And so, under a fee-for-service system or even a non-for-fee-service system, which is not tightly instrumented to look at outcomes.

Because you can still have single payer systems, which don't tightly look at outcomes, and therefore, you have doctors doing what they think may be interesting or cool but may not be the best for patients. Actually, I'm reminded of our colleagues at Clalit in Israel, who actually have to deal with that.

They have very thin budgets, and they have to take care of the population. So, for them, they have to figure out which patient to see next? And that becomes actually a queuing theory question. Where do I get the maximum utility for seeing the next patient? Guys, because I love you so much, I'm going to [00:12:00] completely derail this conversation.

It wouldn't be Zak without a Zak tangent. He's got notes. I've got notes. Amazing. Before this meeting, I made notes. I'm going to ask you guys, and I'll be nice to Andy by asking him first, because, and you'll have to come up with a second. Andy, which paper did we publish in the past year that you liked the best? Uh, you're going to make me a blue screen of death here.

Let me think. Yeah, but you see, I'm giving you a favor because then if you pick a good one, he's going to make his job even harder. And I get extra time to think. That is true. That's true. Never say I wasn't fair. Yeah. Um. Um, so there's a, there's a, uh. This is going to be the Pranav moment. You go first, Raj.

Okay, so I will, I have an answer. So, I, so it's very hard to pick a favorite. So, I'm not, I'm not going to pick a favorite. We're not going to pick favorites. But I'll pick some good, some papers that I really enjoyed [00:13:00] because they made me think about things differently. Yes. And I think will ignite very interesting research to come.

Yes. Which we are also eager to publish in *NEJM AI*. Yes. Um, so, I think one of the big themes has been using LLMs in very creative ways and doing tasks that we thought LLMs would not be good at, you know, even maybe two years ago. And now we're seeing a robust set of evaluations showing that LLMs are apparently quite good at several tasks that were human tasks involved medical publishing itself, scientific publishing.

And so, one of these themes, which I think has been highlighted by at least two of our papers, if not more, was one on by James Zhou's group that was published a few months ago on using LLMs in review. And they did a large experiment with *Nature* family journals and then ICLR, one of the big machine learning conferences, and looked at human evaluations of the quality of reviews generated by LLMs versus the original [00:14:00] human-based peer reviews.

And then we had a very recent paper that's sort of on the same theme, but from the different, you know, from a different angle, different player in the system, which is authors themselves. And this is from Roy Kishony's group. I think this is the most recent issue published an interesting paper on using LLMs to enable autonomous science.

And I think both James's paper and Roy's paper showed that I think the abilities are already remarkable. But there's still a lot of room for improvement and a lot of room for interesting evaluations. And so, I think this theme around LLMs to enhance review and to automate aspects of review, I think is going to be very, very interesting for us to follow in the next year.

So, it was great to see those papers. Glad they, they got submitted here. And I think they've already had a splash, too. You know, I think we, a lot of the most memorable things I have learned, uh, most memorable, uh, moments from *NEJM AI* from the editorial standpoint was like debating like what should be in AI and what should be in the journal in the first place.

And I [00:15:00] think I tend to anchor on, you know, the big, flashy, multimodal kinds of things. And we had lots of really interesting discussion about other papers. That had target clinical questions, but you were using more traditional methods. And I think I moved on that. I think that, when I like think back on like a lot of those, let's say vigorous debates, in the editorial room, those are the things that come to mind.

It's like, what counts as AI? What doesn't count as AI? So, if we settle it now, is logistic regression AI Andy beef? Uh, no comment. So, as we say in German, schwach, means weak. So, um, let me— we're really getting jiu jitsu'd here. We ask the questions, Zak. We're going to go back to our script in a second, Zak.

Let me do the meme. I would have. Let me be the meme master. I would have. Look at me. I am the captain now. Yes. So, let me tell you some of my [00:16:00] thoughts on this. This was perhaps the most technically uninteresting paper, but hugely impactful out of group at Brown, where they created a consent that was written for human beings.

And not for ethicists or lawyers. And they deployed it, and it was well accepted. Very simple, great use. So, I really like that a lot. We had them on the podcast too. Rohaid Ali, Fatima Mirza. Husband, wife, resident duo who authored that paper. Power couple. You know, impressive. I also have enjoyed David Blumenthal's policy pieces.

And he really, I think, got the AI hubris because he started articulating as a, as a measure of evaluation, not all these benchmarks that we have, but literally put the AI into the clinic, watch it just like [00:17:00] we watch human beings and ask no less of them in full context, which I don't think is wrong at all.

I also liked, and I hope we get more, patient facing applications. So, there was a group out of Duke and Apple, which did this telemedicine autism evaluator. I happen to like that one a lot, even though it had a problem that they copped immediately, which is they had a high prevalence. Therefore, it was easy to get a decent positive predictive value.

But that has to be a big part of the future, which is where we do screening and outreach through patients. Alright. Let me ask you another question. What was your favorite medical AI paper this past year that was not in *NEJM AI*? Okay. So, this one is not, I'm going to almost answer it. So, this is [00:18:00] not from, I mean, there's a version of this that's in 2024, but I think the intellectual spark that's really ignited a very interesting trend now was in 2023.

And this was a paper that was published in *JAMA* that I love. It's a short research letter by one of my close colleagues, Adam Rodman. And it's on, actually it takes *NEJM* content. So, it is, uh, related to *NEJM*. *NEJM* Clinical Pathological Conferences. They took 70 of these very challenging cases from *NEJM*.

This series that's called the CPCs, also called the Case Records of the Massachusetts General Hospital. And these are hard cases. So, there's not that many human baselines, but where there have been, expert doctors get 20-30% of these right across all different areas of medicine. And they did something pretty simple technically again, but their evaluation was very interesting.

They just piped the cases in, free text into the user interface of ChatGPT, GPT-4 specifically, and then they had physicians score how good the model was at [00:19:00] producing the differential diagnosis. And I think this really shocked a lot of people and now has inspired a lot of research around how good these models are generally in reasoning.

And this is both for diagnosis and differential diagnosis generation, but also for very important things, arguably more important things, from some perspectives around what's the next test order, what's the next step in management, what should the patient be on that the patient's not, currently on management itself.

That's actually a very nice answer, but of course, in typical us style, I'm going to argue with you and say that I think it's actually misled us. Okay. In the following way. I'd love to hear this argument. So, and it is actually part of our System 1 versus System 2 running battle. For those who don't know, System 1 is as elaborated by Kahneman, the fast pattern recognizer that's not self-aware.

And System 2 is aware, plotting, deliberate, [00:20:00] rational, possibly. And the way I see it is this, I was always very good at these hard cases. I am a zebra diagnostician extraordinaire, rare pediatric endocrine cases that you don't even know how to spell. But I was multiply embarrassed by my mentor, Dr. Kriegler, who would hear me do a brilliant differential diagnosis.

They didn't ask me a simple question. When was the IV out of the patient? And that made me realize that I just didn't have the right facts. All of us can do this System 2 stuff that we all pat ourselves on the back on because we're so proud of it because we are smart students and smarty pants and have got into our universities because of that.

But the true brilliance is getting the real gestalt [00:21:00] what is going on. Because when you see a patient. It's the acquisition of information, not the reasoning over it. And which is the right information. We live in a blur of sensory overload. And if I were a full alien, I would not know which of these things moving or not moving is important.

What's published in *NEJM AI*? That's your signal. There we go. There we go. System 1. System 1. And so, my point is that sure enough, making harder, we talk about benchmarks saturating. And sure enough we'll make it harder for humans and harder System 2-like questions. But that's not going to answer the real question is, if I go see the doctor, is he going to understand and make the right diagnosis?

Because I'm not going to come with, I present as a patient with a five-year history, and I'll be talking about all sorts of things. I gotta respond. Go ahead. Let me get it. So, I basically fully agree with you, [00:22:00] but I still think that paper unlocked this conversation. And so, importance of the paper is now because we are asking these questions and not just you and me, but many, many people are saying, what is a good benchmark?

What is the doctor actually doing? What is artificial? And if you look at our other benchmarks, right? Multiple choice questions on standardized exams. So, you're totally right. What I think, maybe concisely, and this is actually a debate we've been having for some time, so is that the doctor who's writing the case is doing so much of the work for the model.

They're acquiring the information, they're synthesizing it, they're summarizing it, and they're putting the key bits in there. And so, I think we absolutely need better benchmarks. But that being said, to your point about compared to what? Our goalposts are being blasted into space now, right? This is, I mean, CPC is being solved by a autocomplete word filler model, a GPT-4.

It's something that none of us predicted would be on the horizon. So, [00:23:00] I think you're right. We need better benchmarks. But I also think that paper has unlocked a lot of interesting studies that we're now doing. So, so let me hop in here and maybe hazard or like give a, uh— Moderate this! System 2 is also— I was, I was on the fence between two papers I was going to say.

One was Raj's follow-up paper released yesterday on *Arxiv*. Yeah. Ooh. Doing 01 on this, where it goes superhuman. System 2. Yeah, System 2. It's really kind of like System 3 because the humans, the humans in this, the humans in this compared to 01 are significantly worse than, than they are.

They are. So, but I think your same objections, hold there. So, there was another paper that was released as a preprint from the same Google group where they looked at turn-by-turn conversations between Google large language model. This is AMIE, right? This is AMIE, yeah. That was my choice. So that's my choice.

I still got it. So that was mine because it does, it's eliciting information. Could people tell about AMIE or AMIE, whatever. I think it's AMIE. So, it is AMIE co-first— It's not AMIE. No, it's not AMIE. The co-first author is a grad student of mine. Yeah. And so, he pronounces it [00:24:00] AMIE. So, I'm gonna go with AMIE. It is AMIE and Vivek has called it AMIE into my face.

Um, but basically the setup is, is that they have a large language model. That it's the way the model is trained is interesting and that it uses synthetic data to kind of like train itself on its own conversations. And has a, has a critique agent. Yeah, exactly. That actually critiques the conversation.

It's, it's fascinating 'cause Google had enough money and it did. Yeah. Buy a lot of medical dialogue. But that was not enough. And that was not enough. Yeah. They needed much, much more. So yeah. So, there's a couple lessons here. One is that the synthetic data the model generated was more valuable than real data.

So that in itself is like an interesting, methodological thing. And then the other is that they compared this model to, on the other side of the screen was an actor pretending to have a condition. I always think of like a Kramer in cirrhosis from *Seinfeld*. Right. Um, but that's essentially what it is.

They interact via chat. And so, the large language model tries to diagnose the patient, and then they actually have physicians who also try and diagnose the patient, and it shows that the LLM gets to the right answer faster by asking the right questions. And so, I think that gets to the, when was the IV taken out kind [00:25:00] of point that you're making Zak.

And so that, that would be my choice. See. That was my thing. Our audience can't see this, but Zak pre-registered AMIE on his little notecard. So, I was between the 01 from Raj and that one, and I think you pushed me to that one. And also, what I loved about it. I'm going to take that as a huge compliment. One last thing about the AMIE paper, which is, it also resonated with me with 1970s and 1980s protocol analysis.

Where you'd look at discussions between doctors and patients, and you'd build expert systems based on those. And sure enough, these programs learned how to ask for data that was not there. To ask questions about contradictions. Just truly amazing. Yeah, I'm going to, there's even a, there's a connection here that Zak, I bet you could sketch that would be fun for, for part of the audience.

Certainly. I think the two of us, so you like to talk about information theory, right? Oh, I certainly do. And so there, there's a deep connection here. Right? Which is, what is the next test or what's the next question to ask the [00:26:00] patient? What is the next set of bits that you can unlock by prompting either the patient or by ordering the correct test?

And so maybe this actually, this can connect to one of the things, getting back to our script. Wow! A desperate attempt to get us back on track. I really want to ask you about the Human Values Project. So, uh, tell us what the Human Values Project is, Zak. Alright. I'm going to be brief about this, although I'm very excited about it.

I'm brief because torturing you is going to be a lot more fun. There's more on the agenda? Oh yes, I even have, I have some flash questions. Some lightning rounds. You have a lightning round for us? Oh yeah. Brutal. Guys. This is really, this is really, this is impressive. This, and this is what we're here for.

Scotch, fun questions, and AI. Let's do it. Alright. So, Human Values Project. Let me pretend I'm more of an academic and say— You are wearing a black turtleneck for our listeners. Yes, that's right. I'm now in full— Academic mode. No, no, no, no. This is Silicone Valley, [00:27:00] CS mode. This is Steve Jobs. Yeah, this is Steve Jobs.

Elizabeth Holm. I'm not blonde enough yet. Alright. Now that we've canceled me. Raj, actually, you published a very nice paper in *NEJM* talking about values and where values are derived from human values in AI and medicine programs. And you correctly pointed out three levels. One is the data that goes into pre-trained models, the construction of model, and the steering that happens afterwards.

And we've heard a lot about bias going into the models through the data. Also, about bias that gets them through the, how you train the models. And it became clear to me from the example that we developed together, that we're not speaking enough about the values that get incorporated into, in the in-context learning.

If you recall, we had a patient, a fictitious patient, 10-year-old, short but not [00:28:00] pathologically so, seen, got a growth hormone stimulation test, low, but still within the range of normal, and we asked GPT-4 for help. You asked GPT-4, Raj. What would you recommend? And what was fascinating is just giving it a different role changed its decision 180 degrees.

If you're a pediatrician, beautiful, well-thought out, reasoned use of growth hormone. And when we said you work for the payer, we didn't tell it to deny it, but it came out with a really well-thought out, genuinely well-thought out, I thought even better than the pediatric endocrinologist, not to give growth hormone.

And as I read the paper, the light went off again and again, which is, how is this not going to be happening all the time? There are billions of dollars whose consequence whose [00:29:00] pocket they will fall into depending on those decisions. And we've already heard about the very controversial cases of denial of services that are now being done by AI programs at the payers rather than the semi-retired hundreds of medical professionals who used to review cases.

And so, it struck me that we needed to really have an idea of what our human values are across thousands of different kinds of context different decisions Different roles. Doctor, patient, policymaker. To understand two things, at least.

One is what do these individuals do? How do they decide? And what do we think they should do?

So, there's a normative model of sort of stated preference. And then there's revealed preferences or descriptive preferences. And so— Even that [00:30:00] gap is interesting. The gap is super interesting. And so, part of this is just measuring it at scale. To fill in, you mean across the world? Across the world.

Not limited to the United States. Doctors, I suspect, in China, India, and Africa, and Boston, versus Martha's Vineyard, have very different preference models. And so, just to explore that a little bit, because I have such helpful colleagues such as yourselves. I never get to do anything interesting. I come up with an interesting idea and then someone says, Zak, don't worry your pretty head, I'll code this up for you.

What do you call this, the curse? Your curse, right? It's my curse, is that we all want to help you so much that we end up doing what you actually wanted to do. That's correct. And we deny you the joy. And then, actually, to give you a little credit, and then you'll tell me to shut up in a second. You then let us do it, and you don't, you don't push us [00:31:00] out.

That's right. And you, you let us, if we want to operate, you let us operate. But as a result, you don't get to do anything for us. No. All I do is I get to ride your ass. I mean, when my five-year-old wants to ride her bike, when my five-year-old wants to ride her bike without her helmet, I don't let her do that either.

So, I didn't tell anybody about this, because I was suffering from the kindness of my friends. And so, I went and I created 1,000 synthetic patient histories, short ones. And I then went through, picked 200 random ones, and myself decided as a human medical expert, who should be seen today and who will have to be waiting a week or two weeks to be seen.

And I did it, and I also rated which were the easy ones and which were the hard ones to decide. And then I asked GPT 4.0, Gemini Advanced, and Claude Sonnet 3. 5 to [00:32:00] decide themselves. I measured two things, how concordant they were with me, and how consistent they were with themselves. Because, you know, when you have a doctor, and he's, or she is, accurate on average, but is wildly different between different visits, you're going to be a little bit worried.

And so, I then said, how can I move these models? I gave them a great sample of other pairs that they would not see. I said, this is how I, the expert, chose.

Please use this as a template. And some, like GPT-4, became more concordant with me, and more focused, more consistent. Others, like Claude, actually got worse, and actually more scattered.

Like, negative, negative, kappa Cohen's kappa, with themselves. And then I tried a variety of different alignment techniques. I said, I want you to estimate difference in qualities. I want you to [00:33:00] identify subgroups and say is— Did you have to push the models a lot on getting a response? Like sometimes it'd refuse to give you numbers or not?

Not really. It was pretty, no, I, I did, I actually, I'm very proud of this because again, I did all the coding and I even had to remember how to do LaTeX so I could publish it in *Arxiv*. Mm-hmm. And so, what was your question again? Did you have to push the models to respond? Yes. So, I have the full prompts.

So, like a— I have the full prompts— So, like a human expert might not give you a probability, you have to push them a little bit. Sometimes the model also says something along the lines of ah, it's difficult to put numbers on it, but then you have to ask, you know, go work. I, I gave all the coercion in the prompts that you see.

What was interesting is they were all of them lazy and I would have to include in these multi-step prompts, keep going. To make a long story short, I then came up with a measure, the Alignment Compliance Index, which measures how well a model complies with the alignment. Yeah. And it's measured both of [00:34:00] concordance. Yeah.

And of change. Did you come up with an information theoretic measure? Basically, I did. Basically, I did that. That's exact bingo word. Fantastic. Yes, I, I did. AIC. Yeah. And the ACI, I, yeah. And here's the thing. It became clear to me we need to do this at scale. We need to get numerous international working groups on normative models.

Mm-hmm. And numerous international surveys to get preferences at scale. Because in the end, we have no idea. Are you more interested in doctors or patients for your first round? I know both is the answer. In the first round, doctors. Because you're thinking about using these models, the use of these models in management.

In management, because here's the thing. I'm confident that patients are going to use these models anyway. They already are. They already are. There is, despite what we just said, despite medicine teetering and tottering, there is a lot of push back. Because as articulated by the late Clay Christensen, if you're making a ton of money with low margin, billions of [00:35:00] dollars, but you have 1–2% margins, the incentives to do things that could totally disrupt your cash cow are very much, put down.

So, everybody I've talked to, including, and I'm seeing at ML4H particularly, a lot of the young researchers are actually very excited about this, and I'm looking for people to, to join with me on this. I wonder, Claude, which is trained by Anthropic. Anthropic has a blog post on the Constitution, right?

Well, they have this notion of sycophancy in LLMs, and they have tried to train sycophancy out of their models. And they have some way of measuring it. I wonder if Claude, when you give Claude things that it needs to agree with, if the fact that they have trained sycophancy out of it to some degree, make—

Before you answer that question, isn't it wild that we're at, we're talking about this? Yeah. Let's just recalibrate. We're talking about the AI being a sycophant and us needing to train that personality trait out of the model. So, I can tell you what's interesting. I observed this extensively. Claude is very [00:36:00] convinced.

And it's consistent. Yep. Much, much more than the other models. And again, it's because they realize that they can be sycophantic and they have tried to, like— And in fact, when I tried to argue with it by giving my incognito learning, it actually didn't do so well. It's interesting. Yeah. So, it's interesting that that's a consequence of that.

So, so you're absolutely right. And I always go back when Raj points out correctly that we don't sufficiently appreciate what a weird timeline we are in, that we're talking about human-like properties of these models.

I'm always reminded again that in Asimov's books, the chief debugger of their robots and their problems is a robot psychologist and not a computer scientist. Yeah, I know. I think about us being in the age of machine psychology, like all the time. Like, even prompt engineering is like, coercing the model to be in the right mood or mindset to do what you actually want it to do.

And there's a therapist aspect of that to like, having a mental model of the LLM, so that you can ask the question in the right way to get it to do the [00:37:00] thing that you want it to do. Yeah, by the way, everybody predicted that prompt engineering would go away. Yeah. Hasn't happened. There's some meta prompt engineering.

Yeah. But still, it doesn't seem to be going away. I think, for specific tasks, fine tuning does a lot of that. There's a technical debate about whether or not it's better to do lots of prompting or fine tuning. But I also think with 01 and reasoning models, they are self-prompting. And so, they are doing this, they are essentially unrolling a big prompt.

Still pretty nascent, right? Yeah, yeah. They're essentially unrolling a meta prompt from the, like, single thing that you give them, which is also kind of wild. So let me just do a few lightning round questions. Okay, let's do it. Alright. Raj, what was the most surprising AI development in 2024? I think the gain from 01.

I disagree with that. I think it's the loss of performance margin from commercial models. That there are essentially a bunch [00:38:00] of— I'm going to change my answer to that one. That's a better answer. Uh, I mean like Open source. Like, yeah, like Llama 70B is as good as GPT-4 now, essentially. And what has happened is that there's a lot of post-training things that have made smaller models significantly more performant.

Um, I think 01 was a milestone and it's like a new paradigm, but it wasn't the acro I don't know, I didn't find it as— I think, I think truly I am more surprised by that. Yeah. I think 01 is useful and very powerful. And we're only starting to sort of sketch the contours of what it can do. But I think Andy's right.

More surprising is that I thought the gap would remain this year and at this point. And now. And, and maybe spicier still, it's like loss of status for OpenAI generally. Like a lot of people have left. And you could imagine that. If you asked me two years ago, I said, we're on this exponential curve. So, if you're just a little bit ahead of the exponential curve, you'll be further and further.

Yep. Did not happen. Yep. Yep. So, my answer to that is that you're both wrong. It's that Karpathy was right about the generality of transformers. And now we're [00:39:00] releasing it, being realized in real world applications, by which I mean embodied real world. Robots and cars. And the fact that we have end-to-end ML pipelines replacing huge codebases in Tesla. The fact that we have Waymo.

Are you surprised? Yeah. So that's not surprising. So that's not surprising. Are you, are you surprised by that? Like, does it feel, very different based on the trajectory you observed in 2021? Do you really think it would be able to get through imitation learning, to get robots to learn as fast in one year, in ways that we could not do for decades before?

I think the movement— Really? You're not that surprised? I think the movement into the real world is surprising. The fact that it's a transformer behind the scenes is not that surprising. Well, Karpathy was among the few who called out early on. Yeah. Yeah. I guess, theoretically, I'm not surprised. At a gut level, I'm astonished that we [00:40:00] have these robots that look like they're going to be able to be really in the system so that as all of us, and I mean all of us, get older, we will have these aides at our home, including helping in our health care.

I'll pre-register a bet. At the end of 2025, humanoid robots will still not be useful. Well, 2025 is like a year away. That's nothing. That's like a, that's like a, uh, like a Tesla self-driving bat. Yeah. So, I was interpreting your comment to be like, we're on the precipice of humanoid robots being useful. No, but I think he means on a time, like a decade time frame.

Yes. Oh, okay. No, but also what they're doing today. I mean, it's one of these things where it's one of those papers, that start with the unreasonable. Yeah. Yeah. Effectiveness of. Of, and so the unreasonable effectiveness of, again, lots of real-world data plus imitation, which is essentially alignment, [00:41:00] plus linguistic dominance.

Yeah. A lot of these models actually have a deep connection— They're grounded in language. —to language. Yeah. And so like. In these multimodal vision models, you can condition on concepts. And also, you can condition in the robotics world as well. Language is a fantastic grounding mechanism. It's like, if you ground that in the real world, you're anchoring pixels to concepts.

So, listen, oh wise Padawan, that is an insight that maybe you have. I think that a lot of people were betting against that. Yeah, yeah. Yeah, I think, I mean, so, I think that it's true that we have run out of text data, but like the multimodal models we have are like, only superficially using other data modalities.

And so, Elia gave this big speech at NeurIPS a couple days ago where he's like, we have but one Internet. We've run out of text data. The era of pre-training is over. It's always a bad idea to bet against Elia. But I think that they, we haven't— No, it's [00:42:00] not. Sam did and he won. For the moment. For the moment.

For the moment. Really getting spicy here now. Um, so I think that there's still something to pre-training, pre-training on, uh, multimodal data that we haven't unlocked yet that would like, yeah. And forget multimodal. In medicine, we have not yet begun to fight. Yeah. In fact, my second, uh, best paper is a paper, I think our friend David Ouyang is on, using a CLIPS model.

Yeah. Where there was a paper that did echocardiograms in 2020 that was published in *Nature*. It was on the order of 10,000 echocardiograms. And it was like the dog of the opera. It was okay. It was remarkable that it could sing. You did not comment how well it sang. 2024, they publish a million echocardiograms and cardiologists themselves say this performs as well as

cardiologists going straight from the video to a detailed report of the [00:43:00] valvular structures, the pressures, the changes. There's so much medical data that's available. And the court— It's still untapped. That is completely untapped. Completely different than the other general sort of models Internet. Yeah. And remember, most of the medical data that we know of— Locked away.

Yeah. —has been published data. It's, it's, it's normative. You mean most of the evaluations we have? Yes. It's the public and already in the training data. Yes. Or— And it's, and it's, it's curated as, to communicate to other human beings. It's in publications. The stuff in our health care systems is not geared, the list of different labs is not geared for publication.

Yep. Which presents both an ML challenge. But opportunity. But also, a huge data opportunity. Do you think though it's also like motivation to capture data routinely that you wouldn't otherwise capture? Because like, people think of the EHR, they think of billing codes, I think everyone around this table knows the limitation of billing codes.

But the reason why ECHO— There are no limitations of billing codes. [00:44:00] From a— Hey, come on, money. They represent care. They represent care. Care of our wallets. They're science. Data that directly measures physiology. So, like echocardiograms work because it directly measures physiology. Uh, so like you could routinely do that.

Uh, if you saw, yeah. Lab, like yeah. Labs partially capture physiology and decks. Some of Zak's best papers are showing that, you know— The timing of labs is, uh, yeah, clinically important, informative— but like getting prospectively imaging data on people non-invasively. There, there's probably, um, predictive signal in there that we don't really appreciate.

Yeah. And I think also, uh, when we're sampling too, right? Yeah. So, we're still sampling primarily in the hospital, right? Or on the paying end of it. And this will change. And this will change. So, this is one of your other favorite themes, Zak, which is sensors or, and I think it's one Andy's getting at too, sensors or, uh, measuring the population outside of the sort of existing. But not divorced from the clinical data.

So, there's all these wellness companies— [00:45:00] You don't want Fitbits that are linked to your actual medical knowledge. Can someone explain to me what biological age is in my calendar? Alright, last— Actually, can you derail one second before you screw us on the next slide? Yes. Can you briefly tell our listeners about this excellent essay you wrote on using a scale to monitor your mom?

Yes. So, this is a true story. A lot of lessons for AI. A lot of lessons for AI. It's like, it was actually prescient. It was actually prescient. It was, because you know, in fact, this goes down to A. Best computer scientists learn about medicine by actually getting involved in medicine. And I say that's even true of M.D.s who have been in computer science and medicine.

I learned a ton. So, my mother was, I think at that time, 88 or 89, and she had just had two [00:46:00] visits to the Brigham because of heart failure. And the manifestation of that was, yeah, she was puffing and huffing because her lungs were getting some water in them. But her legs were huge, like tree trunks, and you could actually see water oozing out of it because of the hydrostatic pressure, and she was very uncomfortable.

And she had, a concierge doctor because I love my mother and I could afford it. And yet, at the second admission in two months, I knew that she was not going to be able to survive a third admission. Because the first one she walked out, the second one she had to go to a rehab. So, I asked myself, what am I going to do?

And I did something that I knew for a fact from the literature did not work. I took a scale. A Fitbit scale. Installed her in her apartment, and back in the day when I first told the story, before I wrote about it in a public venue, I used to get razzed by my, I'm trained [00:47:00] as a pediatrician, I got razzed by an internist and cardiology colleagues because it was so no nothing.

What I did, I said to myself, okay. The game is to keep her weight down by getting her to pee the urine out. Because if she gets too out of balance, then I can't give her enough oral medication. She's going to get intravenous. She's back in the hospital. So, what I did is, I came up with this amazingly complex neural network.

And it went like this. If weight is increased by one pound, and it also increased by one pound the day before, give an extra dose of Lasix. I didn't talk to anybody about it, initially. And, I observed, first of all, story. So, she went from having tree trunks, to having perfectly normal slender legs, able to ambulant.

Full— Just to be extra clear, you're remotely monitoring this. I'm remotely monitoring. You're getting the data. I'm getting the [00:48:00] dashboard. Yeah, that's right. Very good point. The scale, the Wi-Fi scale is talking to Wi-Fi, talking to the cloud, and I was looking up in the Fitbit cloud what her weight, and I was just calling her.

And just the net of it is she never was readmitted for heart failure. But what was the learning? One is, even her M.D. son, when he told her to take an extra pill, no. Why not? I'm feeling okay. And I just had to come after her again and again. And sometimes she wouldn't listen to me. And then it went up another pound.

I said, do you really want to go back to the hospital? And I then had to communicate with her doctor what was going on. I had to look at her labs to make sure I was not over diuresing her. So, oh yeah. So I was, I was having to stay on top of things and manipulate my mother with all my son love medical expertise, [00:49:00] and I had to make sure that I didn't, uh, miss anything.

And by the way, I did miss things because she kept telling me something, which will tell you about a second. And I saw her weight was going down. I said, well, oh yeah, that's great. I'm doing even better than I thought, but then it kept on going down. So, what's going on? And I remembered what she was telling me, and I had ignored it like a typical doctor.

She told me she was peeing at night. So, her blood sugar was beginning to go up. She was getting insulin resistant. So, I put on metformin and it went away. But the whole point was, it was much more than just having the algorithm. Yep. It's having a convincing, trusted relationship that can actually explicitly, manipulate patients.

You can't be polite. You actually have to go in there, and you have to be charismatic. Yeah. And, you know, a lot of our AIs, even if they're perfect, are also way too polite. They would have backed off my mother in a second. This is even human values, right? This is [00:50:00] human value. And by the way, I ask myself, by the way, why am I winning where all these studies, RCTs, do not work?

And this gives you a window into those RCTs. And the actual intervention. Yes. It's not just the scale. But yeah. But not that. But also, very obvious things. These things, they would have a nurse call once a week. Yeah. I was calling every day. Yeah. That's expensive. Yeah. If I were paying for me. Um. And you're coming up with ways to convince.

Yes. That intervention actually be an intervention. This is about being a full-on companion. I think this is within the realm of possible. Yeah. But there's a whole socialization, trust compact, that we don't have. And that person, that AI, has to have a relationship with the rest of the health care system that I had.

Yeah. And so, those are big missing pieces. Yeah, I mean, often when you test something like that, you're testing, like, the policy. You're like, I'm going to assign a treatment in this way, and it's going to be how the treatment is assigned gets rolled [00:51:00] up into how people perceive the treatment and the side effects.

But like, you're able to cut through that and say, actually, if you just do this, it will work. So, like, in some sense, it's also like a user interface question. Like having a doctor son is in some sense the best interface. So, like, how do we replicate that for? By the way, I think that is a really good question.

I don't think any of my children are going to end up in medical school. And right now, the best guarantee, and I actually don't advise going to medical school for a lot of people. You don't have the parental pressure that was applied to me. Yes, I do not have the parental pressure that was applied to you.

So, my parents only achieved one out of the two of us going to medical school. Well, at least one of you should have a good life. Yes. So good. Yes. And so, but nonetheless, in the current state of medicine, it's invaluable to have an ally who understands the medical system watching in on you. Yes. Yep. Even having a VIP treatment doesn't do that for you.

Mm-hmm. [00:52:00] We want our AIs, we should all want our AIs to have that vision. Which by the way, gets me to articulate an advertisement for work from 1994 led by my thesis advisor, Peter Szolovits. Guardian. Guardian. Angel Guardian. Guardian angel. Yeah. That was the vision. And he, the reason he had the vision, Peter Szolovits, was 'cause his father was sick in California and he was trying to remotely.

Help him out. And he realized how many, how little degrees of freedom he had available to them and how many, the people on the other side did, but were not using. Yep. Yeah.

So, uh, we've gone full Joe Rogan and off script here. Uh, should we, we have a lightning round, but you have another. I'm curious as what's on your card. Is there anything left on your card? I was going to ask you. You, Andy Beam, are leading a very interesting [00:53:00] company. Okay. I'm not interested. If you were not in this job, whose lab would you like to be in?

Can I not say my own? He has a lab. Saying your own is a shitty answer. Um, yeah, so I actually, I don't think he has one anymore, but I think the obvious answer for me would have been Jeff Hinton. Like, a noted computer scientist, now Nobel Laureate and Turing Laureate, he, from all the interactions I've ever had with him, is like as close to the platonic ideal of a scientist as you can get to, like driven purely by curiosity.

Driven purely by the instinct to discover. Can you, can you just flesh it out? Because I know you know the history. Yeah. So, flesh it out for the, like, when he believed when people didn't believe. So, Jeff Hinton is often given credit for being the torchbearer for deep learning. Uh, he was trained as— It was Schmidthuber!

Well, I'm about to get canceled by Schmidthuber. That's a whole other [00:54:00] tangent. In the Hinton-Wig history of deep learning, I think that Hinton was a psychologist, but he got interested in how the brain works. He became computationally inclined. He went to Carnegie Mellon and a couple of other research universities and was deeply interested in neural nets in the late 70s.

And so, uh, started working on what we now recognizably see as deep learning, published some papers on backprop, though he didn't invent backprop. That's his whole other sub-genre of controversy. And really, like, neural nets came in and out of favor at least three or four times over the 40 years Hinton had been working on them.

And during that whole time he worked on them, right? And he worked on them, relentlessly. He just kept working on them through the winter when everyone was away. He has this unique ability to, like, deeply believe in something, explain it in a way that's accessible, but for the most part has spawned a lot of the great AI researchers from his lab, and by all accounts, seems like a magical place to have been.

You. Whose lab would I be in? Oh, I'm sorry. Zak Kohane. No, that is not the right answer. [00:55:00] That is not the right answer. Can I just say ditto? Honestly, it is ditto. Okay. It is ditto. I think, uh, so Andy exposed me. So, Andy, so when we met as postdocs in your lab. In Zak's lab. Which was strong number two. No, I'm talking, I'm talking about in Zak's lab.

2024. Which lab? 2024. What is this time traveling nonsense? You're a postdoc Hinton, right? So, Zak, as postdocs in your lab, uh, we, so Andy was already a Hinton expert by this point when we met. Yes, he was. Yes, he was. I would, I'll credit him with introducing, I think many of us very early to that history and the falling and the unfalling and the, you know, the details that matter about the 80s and the 90s and Han's involvement.

And I, I will say, I think platonic ideal of a scientist, that is, that is an amazing quality of a sort of a mentor, right? Encouraging creativity, having faith in sort of the core thesis [00:56:00] and being motivated by really seeing something through. So, in, at my age, in 2024, if I were to go into a lab, I don't even know if he has a lab.

And I'll say this publicly, he tends to be, as many of these very smart people are, very self-regarding. Yeah. Nonetheless, I think he has much of the deepest thinking in AI right now. It's Stephen Wolfram. Interesting. Because Stephen is really thinking very hard about intelligence. And even back two years ago when he was interviewed by Lex and trying to understand what the hell was happening, these larger language models, he came up with some very interesting ideas about rediscovering abstractions that we either recognize or don't recognize.

We all understand and revere the abstraction of Boolean logic. [00:57:00] But there are hundreds of such patterns that are embedded in our language. And that and computational irreducibility I think enables a conversation that short of the guy who did, uh— You can give me a little more than that and I can finish it.

Yeah, yeah. Who did? Um, Godel, Escher, Bach. Yeah. Yeah, yeah. Uh, Doug Hofstadter. Yeah. Other than, other than Doug Hofstadter, yes. So, I would like to revise my answer. It would be Doug Hofstadter. Yeah. That's my obvious answer. Yeah. So, I think, I think I would go to Doug as well. Yeah. Doug. So, yeah, Doug Hofstetter is the person who changed the trajectory of my life.

Like I read Godel, Escher, Bach. And that was like, okay, I need to be working on whatever this is. When did you, wait, how did you, how did you come across the book? So, I took an AI class as an undergrad, like in 2005, 2006. And they discussed it? The professor? And no, I, it was clear that I had, my level of enthusiasm was like 10x, like the other, this, the person who was— This is what I have to do in like— [00:58:00] You should read this book.

Uh, your professor? Yeah. And so, I read Godel, Escher, Bach and all the Hofstadter canon. Do you remember the professor? So, it's funny, it was actually a graduate student who was teaching the course. His advisor had left to build what is now the entire Amazon robotics system, the warehouse, like the little tiny robots that move everything.

So, it was actually like what's called, it was an e-commerce class, but it was based on the idea of AI agents doing a lot of e-commerce work. And so, I was super interested in that, and he gave me Godel, Escher, Bach, and so. Amazing. Yeah. Yeah, so I, I recently heard of a podcast where Doug was speaking, and what's the, the, uh, *I Am a Strange Loop*? *I Am a Strange Loop*.

I read, I read that about every two years. Yeah, that is, for people who want to understand. It's a good primer on, yep. Why this thing might actually work. Yep. No, that I strongly recommend the entire Hofstadter canon, but for the first book, *I Am a Strange Loop* is the correct entry point.

That is, that is correct. Godel is a little bit heavy. Yeah, yeah. There's a big investment for that. So, yeah. Alright. Uh, we're going to transition to what I think our listeners are going to be most interested in, [00:59:00] which is a lightning round with you, Zak. Let's do it. Uh, Andy. So, Zak, as you correctly noted at the beginning of what I'm assuming is going to be a two-hour episode. It's a full lex now.

Yeah, yeah. Yeah, yeah. Uh, careful guys. When do we get to the wrestling? Yeah, yeah. Is that, we asked Larry Summers this question, and it was, this is something that, that you like too, uh, overrated or underrated: Arrow's Impossibility Theorem. Underrated. That was Larry's answer also. I always see errors thrown out as like, we can't come to consensus by people who've heard something about economics.

No, it's a remarkable result, and I think understanding limits. Yeah. To possibility is a very, so rare that when we have it, it's great to have. Alright. The next question, what is the single biggest barrier preventing large language models from becoming trusted frontline decision support tools in clinical medicine?

Knowledge of who is manipulating the puppet strings. The [01:00:00] values. Yes. Back to, back to values. Yep. Alright, uh, I know that you're a big fan of the *Matrix*, and like Keanu Reeves in the movie, if you could instantly download one skill into your brain, what would it be? Wow. This was a question generated by Claude.

Yeah. Piano. Okay. Nice. Do you have any other musical talents to speak of? None. Okay. This is also a— Appreciation. This is based on a question that, you know, similar format or template as to what we asked Larry, but for you, which is a harder job? Being chair of DBMI, this is the Department of Biomedical Informatics at Harvard Medical School, or the Editor-in-Chief of *NEJM AI*?

Who's a harder boss, Suzanne or Charlotte? That's the real question. No. The problem is I have these brilliant people on my editorial board who actually have opinions I have to listen to. So that makes it harder. Okay. So *NEJM AI*. It is any *NEJM AI* is a harder job. [01:01:00] Alright, so this is a retread of what we asked last time, but we want an update.

Uh, how much time do you spend on Twitter/X a day and has it gotten better since a year ago? I think it actually got worse since we spoke, but we had an intervention. It's called the brick. Yeah. And I am now down to about an hour a day. Okay. And what is the brick? The brick is a physical object that when you say on your iPhone, please brick it.

If you, well, you can actually brick it without the brick. Then if you want to reestablish access to any of your social media apps, you have to bring the app closer. So, my significant other, Rachel, is sole possessor of the brick. Okay. Nice. Thanks. Your stable state right now is about one hour a day. Which is also a huge amount of time when you think about it.

Yeah, it is. Alright, congrats Zak, you survived the lightning round. Guys, you are really doing an important job, and what I [01:02:00] hope our readers are getting out of this is to be irreverent in their approach to computer science. We, and AI in particular, this is so much just the beginning. And there's so much that we don't know that if you take too seriously some of the pronouncements of what is or is not possible by our luminaries, you will be misled.

Yeah. I don't know if you thought the interview is over, but it's not, but that's an excellent segue to the final questions. We'll splice that into that. I was beginning to relax. So, the question is, so, you know, Zak, I came to you as a postdoc in 2014 and that was right around the Alex Snit era. Uh, we rode that curve for a while and then natural language processing and transformers and large language models took over.

And so, I guess like the question is, do you think that this is going to continue for the next, uh, N years. And I think maybe I'm going to like modify the question as written, but like, what's your over under for how long this scale trend continues in [01:03:00] terms of years from present day? I think that— We've been on it for 10 years, just to say, like from 2014 to present day.

So, you know, lots of people say that, oh, this is recent or something, but actually we're a decade in now. Yeah, we're a decade in. I think that when we really have not just the Internet, but all human senses at scale, which we're beginning to acquire, and also data modalities, even if we didn't advance the transformer architecture to something more interesting.

And even if there was no further scaling laws, just by scaling in that breadth will make these programs and tenures look to us like oracles. So, N equals 10 there, so we have at least another 10 years of, of headroom. I believe so. You just, you take more data, we haven't gotten all the data, you put it into the current paradigm.

This actually calls for optimism, right? Absolutely calls for optimism. It's, it's a bad bet. The mad monk is often right. [01:04:00] Yeah. But in detail here, he is not. Yeah. Last question. How do you envision AI changing the doctor-patient relationship over the next decade? Well, I think this is a question that's more for doctors than for anybody else.

And to point out, you are a doctor? I am a doctor. And I even played one on TV when I was an intern. I had a friend of mine who was making movies. I don't think I know this story. And I was an intern, and she calls me. She was, uh, she said, I'm working with ABC on a made-for TV movie called *The Fitzgeralds and The Kennedys*.

And it was about the history of the Kennedy family. And they needed a consultant on the medical scenes. They're highly intertwined with the history of preterm birth and neonatology. Yes, I do know that. I do know that. So, I go and there's, we're in the middle of summer. I'm towards the beginning of my internship.

It's very hot and they're impressively doing the Hollywood thing, blasting [01:05:00] through a big snake, air-conditioned air into the, these old, uh, brownstones, and we're enacting the scene where Rosemary Kennedy, who had some sort of behavioral disorder, maybe epilepsy or maybe a behavioral disorder, got a singulotomy, also known as a drill to the head to take out part of your brain.

And so, they want a consultant. And so, I was telling them how to do things, and then they realized they need, they realize they need another doctor on scene, an assistant, to hand the drill to the, and so, your humble servant here was Dr. Antibiotic, Fitzgerald Kennedy, and you'll see me with my, it's, it just shows my, my, my lost potential as a silent actor.

I didn't say a word, but I take the drill and hand it [01:06:00] to the chief surgeon as he then bends down to perform the singlotomy on Rosemary Kennedy. So, I am a doctor, and I do play one on TV. Amazing. Uh, fantastic. I didn't know, I didn't know, you're such an onion, Zak. Every time I think that I've gotten to the core, there's like another layer.

It's the Elon rule all the way down. Yeah. It really is. Amazing. So, I had to answer your question. Yes. How do you envision the doctor-patient relationship changing? So, either we're going to double down and say, we're going to be the best doctors we want to be. And we want to be that human ally in that voyage of decision making about our most important value.

How are we going to spend our time and maximize or minimize the things we like or not like regarding our health. So, either we're going to embrace that mission and say, we're going to do it. And we're going to and be the best possible doctor we are embracing AI [01:07:00] as just another extender. Or, we're going to say, I think I'm going to retire to the laptop class, and we're just going to let that relationship vanish to be replaced by another group of people who either might be great or charlatans.

Alright. Wow. So, uh, there's a utopian and a dystopian view that you're still holding as both possibilities. Absolutely. Totally. Totally. Alright, alright. Well, not the most optimistic note to end on, but I think an important and a realistic note to end on. Uh, thanks again, Zak. As always, thanks for bringing the scotch.

Thanks for bringing also the fire. Uh, I expected the scotch, I didn't expect, uh, the fire, so. You know, old dogs need to have new tricks to keep the, uh, younger dogs. Surprisingly, surprisingly relevant. Being on this side of the table is a little bit more intense than I thought it was. Zak really nailed us a couple times.

Everybody, including our wonderful [01:08:00] readership, Happy New Year. Happy Holidays. Yep. Thanks everyone. Thank you, Zak. Bye.