AI Grand Round Podcast #6
05.17.23
Google's Exploration of Large Language Models in Medicine

[00:00:00] And so the immediate questions that come to mind for me around this kind of powerful technology was in the much more challenging setting of healthcare where if a language model makes a mistake or makes an error, there's a sort of much more perceptible risk or harm than in some other context, for example, in in creative applications and other things.

[00:00:24] So one of the scientific questions I think that arises in that moment when the technology is coming to fruition is to start asking the extent to which clinical knowledge and medically important information is actually encoded in these systems to begin with, and to start asking scientific questions around how to best measure that, but also how to begin to put metrics around it and maybe even then optimize and develop it. So at the outset of a research field, we try to make contributions that are generally useful and thoughtful, aligned [00:01:00] with the values of the practice of medicine, and of what matters to patients and people.

[00:01:07] That was Dr. Alan Karthikesalingam of Google, describing his team's efforts to understand how well large language models encode medical knowledge. Welcome to another episode of NEJM AI Grand Rounds. I'm Raj Manrai, and I'm with my co-host Andy Beam. Today we're thrilled to bring you our conversation with Dr.

[00:01:25] Alan Karthikesalingam and Vivek Natarajan, who are both at Google. Alan is a physician and scientist, and Vivek is an AI researcher and they've really been on the leading edge of carefully applying and testing the capabilities of machine learning models and especially large language models of late in medicine.

[00:01:42] ChatGPT from OpenAI is probably the most widely known example of one of these large language models, and we talked a lot about ChatGPT and GPT4 in a previous episode with Peter Lee of Microsoft. On today's episode with Alan and Vivek, we explore how these models are developed and how they're carefully evaluated [00:02:00] for their clinical capabilities.

[00:02:01] Andy, we've seen the headlines about AI doing well on medical licensing exam practice questions, but I think as Alan and Vivek articulate, this progress opens up more research directions than it closes. Overall, this was an informative conversation about a very fast moving field. I totally agree, Raj and I really enjoyed the breadth of topics that we touched on in this conversation, including things like large language models, medical testing, ethics, and alignment.

[00:02:27] As you know, Raj, this conversation touched on some of my own pet projects, like medical question answering, and so it was a lot of fun to chat with Alan and Vivek about their groundbreaking work on this problem. One of the highlights of the conversation for me was learning about the large language models performance on step one style practice questions, which are questions that are used to test med students' clinical knowledge.

[00:02:47] This model achieved a remarkable 70% accuracy, significantly surpassing previous models that were limited to only 40 to 50% accuracy. It's worth noting that since we recorded this conversation, Vivek and Alan's team have released an updated model [00:03:00] called Med-PaLM 2 that now scores over 80%. We also touched on the role of public benchmarks in accelerating progress in medical AI and how their existence contributed to the development of Med-PaLM.

[00:03:10] Um, I really think that their work also sets a new gold standard for the evaluation of large language models for clinical applications. And with that, we're happy to bring you Alan and Vivek on the next episode of NEJM AI Grand Rounds. The NEJM AI Grand Rounds podcast is sponsored by Microsoft and Viz.Ai.

[00:03:29] We thank them for their support.

[00:03:34] Well, Alan and Vivek, welcome to AI Grand Rounds. We're very excited to have you here. Excited to be here. Thanks. Likewise. So, Alan, we like to start with some intro material and learn a little bit about our guests before we dive into what you've been working on. So maybe you could walk us through your career, how you got interested in medicine and what led you to this intersection of artificial intelligence and medicine.

[00:03:58] Sure, Andy. So [00:04:00] I guess getting interested in medicine is probably, maybe a slightly corny but true story that many people say when they go to medical school, which is, I've always found that it was this amazing combination of science and humanities. And I was always very drawn to the

idea that it's a great way to spend your life, is to sort of try and make the lives of other people better.

[00:04:20] And that was really cemented for me by, you know, as a teenager, trying to explore the, explore it a bit. I did work as a theater porter in operating theaters in the hospitals near me. And because I'm not the world's strongest person, I maybe wasn't the world's best porter, but they did used to let me hang around and ask questions of all the, like nurses and doctors and procedures that were going on.

[00:04:41] And I just was immediately hooked. I thought it was the most amazing environment and it's the most incredible things were happening. And then the thing that really sealed the deal was my parents who are doctors telling me not to do it, which obviously to any self-respecting teenager is like a red rag to a bull.

[00:04:58] So I was then really fortunate. [00:05:00] I studied medicine, uh, medical sciences at Cambridge and went into surgery mostly because I could immediately see the benefits. Like I've, I've always been motivated by patient outcomes. And in surgery that was immediately tangible. I could see the benefits really quickly within surgery, I then went into vascular surgery because that was even more the case.

[00:05:20] Basically, you know, limb saving, life-saving interventions. And the other thing that grabbed my attention was the role of technology in that particular specialty. That was also simultaneously what ignited my interest in research. Um, it was basically because doing these high risk procedures and in being involved in the care of so many critically ill people, you quickly become aware of things that can be done better.

[00:05:46] And you also quickly start to see, I think as a practicing physician and surgeon that sometimes are things that harm people or that don't make people's outcomes as good as they can be. Are repeating things and when things repeat, they show up [00:06:00] in data and you can use statistics and the scientific method to directly address that.

[00:06:07] And then you can go from bedside to bench and then find solutions or hypothesize about ways to fix things and then go back to the bedside again. And so I was really fortunate. I then came to London for my surgical training, was able to do my own PhD, which was funded in a, a program that the UK runs called the N I H R, which offers integrated training and worked with some amazing mentors.

[00:06:31] After completing my own PhD, I then ran a lab and had my own PhD students. And we were doing a combination of work with both with medical records type of work and outcomes research, trying to reconfigure high risk surgical care, but also we were at. And in vascular surgery, in particular devices that are used to treat aneurysms and occlusive arterial disease.

[00:06:55] And in both of these areas, in around kind of 2014, 2015, 2016, it became [00:07:00] apparent to me that the sort of statistical approaches that I had learned and that my own PhD students were also developing and doing were one really useful tool in the toolbox. But I was also becoming increasingly aware that most of the most amazing research I was seeing around me was coming from collaboration with completely different disciplines.

[00:07:19] And in particular, I started to get really interested in the ability to work with engineers and product managers and essentially this burgeoning field of digital health. Allied to what at the time was, you know, this awakening of deep learning. Essentially. That's what led me to DeepMind. DeepMind was in London at the time, and I approached academics at DeepMind about collaborations in my particular field of interest in medicine, and was met with the most amazing responses about the actual potential of that field of research way beyond that to the whole of healthcare really.

[00:07:56] And I realized that at the time, you know, artificial intelligence research was a [00:08:00] bit like electricity. It was this kind of foundational technology that could be really transformative far beyond what I had been thinking about from within my research program. So anyway, I was, I was very fortunate to then spend a year at DeepMind that year, converted me to being, uh, from sort of a, a practicing clinician who was spending one year in technology to sort of the other way around, someone who wanted to be a clinician inside DeepMind.

[00:08:25] Now Google where there's just this incredible ability to work at scale with product managers, engineers, machine learning, research scientists. In an environment where there's experience in delivering real products at scale to the world and do that as the sort of first violin and have the second violin be my clinical practice, which I, clinical academia, which I, I keep up with.

[00:08:47] But it's, um, the primary area for me now has been, uh, the last kind of seven years at, at Google, which has been fantastic. Awesome. Could I ask a quick follow up there? If I think about the things that historically DeepMind has [00:09:00] gotten excited about, it's been, I would say, grand challenges. So

protein folding, go nuclear fusion, what was it in your conversation, your interaction with them that got them so excited about healthcare?

[00:09:11] Oh, I think definitely wasn't me. Um, I think at the time it was a long time, you know, it was kind of just at the time when there were the first explorations of supervised learning beyond ImageNet. It was that kind of era, and it certainly wasn't me that was proposing any of these things as a grand challenge.

[00:09:32] It was amazing scientists like Olaf Ronneberger, you know, who had had just joined at that time and had made this amazing foundational discovery of the unit. And so there were, there were these magical conversations happening at the time around how, what was progress at that stage with convolutional neural networks in quite low resolution natural images of the likes of ImageNet.

[00:09:54] There was some really foundational scientific questions then about in socially meaningful contexts like [00:10:00] medicine, but when you think about how much more complex medical images are, not only that they're 3D and volumetric, but also just computationally, how much more challenging they are to actually find the identifying features of disease, how approaches like segmentation might play a role and how to actually go about that from a machine learning perspective.

[00:10:17] I think at that time it was its own grand challenge. It's, it's difficult to look back at that now because of course there's been an enormous wave of progress in there that perhaps now it's not quite so surprising. But at the time, that was before any real research had been published, applying deep learning to any kind of medical image.

[00:10:34] Yeah, it's funny how far away five years ago feels at this point. Right? So at one point that seems like transformative now it almost seems like ancient history. So there, there's a lot there that I'd like to revisit later, but I think we'll stop there and I'll throw it over to Raj. Thank you Alan. Um, just echoing Andy, I'm delighted to have you both on, on AI grand rounds. Vivek, we'd love to hear about your background too. I'm really curious in particular [00:11:00] about how you first got interested in artificial intelligence broadly, and also about what experiences led you to start tackling medical AI projects. Yeah, firstly I am delighted to be here and talking research and medical AI with two of my favorite researchers in the field, and I'm even more delighted to be doing this with my dearest friend, colleague, and mentor Alan.

[00:11:22] I grew up in India, I think for most kids back then in the nineties, your parents either want you to be a doctor or an engineer. My parents were more like, you know, you know you need to go into medicine, but I just could not bring myself to memorize all the biology textbooks that you had to do if you want to like track the medical insurance examinations in India.

[00:11:41] So I ended up picking engineering and disappointing my parents along the way. Back then, most students don't end up selecting their specialization for engineering based on any kind of interest or something like that. It's more like if you're ranked on the top hundred on the entrance examinations, you end up picking electrical engineering, the next hundred picks, computer science, the next hundred [00:12:00] picks, mechanical engineering and so on and so forth.

[00:12:02] And so yeah, you just pretty much end up following the herd. And for me it was kind of the same. Uh, I ended up picking electronics and electrical engineering. While the coursework was super interesting, it had topics like semiconductors and single processing, like really foundational topics. It did not involve any machine learning or AI, but I think I was super fortunate to be doing my undergrad at a time when, you know, massive open online courses were becoming a thing.

[00:12:27] And so it was one fine evening. Uh, I was at the internet lab in my institution and I randomly bumped into one of these lectures from Professor Yassir Abdul Mustafa at Caltech, uh, on learning from Data. And I did that on YouTube and I was absolutely hooked on that topic. And I remember spending that entire semester using every bit of data bandwidth that I could get hold of to download lecture videos from that course and from Professor Andrew Ng's machine learning course.

[00:12:55] It's kind of interesting to reflect also on how far the internet infrastructure has [00:13:00] evolved in the past decade in India. Now I think it's among the best in the world, but digressions aside, I can firmly say that I am a product of the MOOC revolution. If MOOC's were not a thing, I don't think I would be doing machine learning and AI today.

[00:13:12] Probably something very, very different. So yeah, that got me introduced into the topic. And so when I came over to UD Austin for grad school, I tried to take as many machine learning and AI courses as possible. But then UD Austin is not like Stanford where if there's a paper, uh, out on archive, three months later there is a course.

[00:13:29] UD Austin was not like that. And so even back in 20 14, 20 15 when deep learning was, I think fairly prominent, there weren't any courses, but I got good grounding in old school AI topics like probabilistic graphical models and reinforcement learning without any of the deep aspects. And the professors over there had been in the field for like, you know, 20 odd years, 30 odd years

[00:13:47] so they had like a very good historical perspective of how AI the field had evolved including, uh, the AI winter in the nineties. And so I wouldn't say they were jaded, but they were like more pragmatic and less trying to hype up the technologies. [00:14:00] And so that kind of always stuck with me. But I think for me the real big breakthrough was when I finished my masters and I fortunately ended up at Facebook AI research.

[00:14:10] That was when I think fair was like really taking off. It was just a year old. And I think one of the best parts about the modern deep learning and AI revolution is that you did not have to be an expert with a PhD or like have these many years of experience to participate in it. Just to give you an example, I think some of the biggest names in the field today, such as, you know, Soumith Chintala, Aditya Ramesh created the DALL- E models at OpenAI or Alec Radford was behind many of the GPT models and much more.

[00:14:36] They don't have a PhD. I think Aditya definitely has a bachelors degree. So the barrier to entry to this field at least was low back then. I'm not sure that is true today, which is something we can discuss later if we had time. But I think that was good. And so it, for people like me, all I had to show was like a willingness to learn and I could like, you know, come in and work and contribute to research.

[00:14:56] And so at fair I got to work in a bunch of different [00:15:00] areas, uh, speech recognition. NLP vision and robotics. And while back then it was not like today where every field is literally using a transformer variant. There were still many common themes around the model architecture, the way you learn these models, the underlying frameworks, the engineering.

[00:15:17] There were a lot of common themes and these were repeatedly getting used across these domains and seemingly, you know, very different problems. I think the best part was it all worked. And so I saw these models repeatedly, like reaching state of the art performance on research benchmarks, breaking through performance ceilings, not seen in like maybe decades, but also getting shipped to production with, you know, millions of users improving, like lifting key metrics in ways not imagined, and also enabling magical new experiences.

[00:15:42] And so after a few years at Fair it, it was obvious to me that I think this AI thing works, although it did not necessarily have the convergence that it has today. And so I was generally thinking, where is this going to have the most impact in the next decade and beyond? And to me it felt like that was medicine primarily because there was a [00:16:00] couple of really interesting papers that came out around that point of time.

[00:16:03] One was from Andre Esteva and others at Stanford in Nature on Skin Cancer detection. And then I think it was Google's own work in diabetic retinopathy. Around the same time, there were a few incidents in my family where it felt like if people had access to better and timely care, the outcomes would've been far, far different.

[00:16:21] To me, it felt like, I mean, if we really want to scale up world-class healthcare to everyone, then AI is our best bet. And so I was incredibly motivated to work at the intersection of AI and medicine. And fortunately at the same time, Greg Carado, Dale Webster, Lily Peng and others were spinning up Google Health with researchers such as Alan from DeepMind and others from Google Brain, and I got the opportunity to come in and I did so without any hesitation.

[00:16:45] And I would say it's been a blast getting to work with people like Alan in an extremely smart, welcoming, diverse, and an interdisciplinary team on challenging, but, uh, meaningful problems, as you would all appreciate. So, yeah, I would say, [00:17:00] uh, looking back, it's been a bit of a diverse pathway. I, when I started off, I did not know I would be working on AI, let alone medical AI, but I'm super glad to be here.

[00:17:09] Great. Yeah. Thank you. Vivek. I think it's, it's fascinating that you, you both took such different paths but have arrived both at, at medical AI. I'll also note that they're different paths, but it sounds like a common thread is disappointing your parents and following a different path that now seems predictive of success in medical AI.

[00:17:27] So I wanna transition to your research now, and the place I wanna start is with your recent paper on Med-PaLM. I was scrolling Twitter a few weeks ago, and I saw this great thread by Vivek that announced the paper and it really caught my attention. I have the first tweet copied here. It's our LLMs, our building on Flan-paLM reached SODA on multiple medical question answering datasets, including 67.6% on MED QA, USMLE, greater than 17% over prior work. So, you know, there are a lot of important acronyms in that sentence.

[00:18:00] One that's gonna be very familiar to many of our listeners. It's of course the USMLE, the United States Medical Licensing Exam.

[00:18:07] Maybe starting with one of the other terms that may be a little less well known, LLMs, uh, but which is in the title of your paper, Large Language Models Encode Clinical Knowledge. I wanna start with a question for Alan. Could you maybe just give us an overview first of this paper, this project, how you get started with it, and then what the major results are of the paper?

[00:18:27] I'm also always fascinated with the process of task selection in these types of medical AI papers and how important it is to understand, uh, what the task was to frame the results. So if you could also tell us about what specific tasks you used to test your models and then how you ended up selecting those particular tasks.

[00:18:46] Sure thing. Yeah. As I'm sure Vivek will describe much more expertly than I would, uh, ever be able to. I think one of the kind of ingredients here was that in the AI field in general and in particular at Google, there had had been [00:19:00] some, uh, really outstanding progress in the field of Large Language Models, and we were increasingly seeing that with scale of these models was coming, I think was being published as sort of emergent properties and really kind of surprising new capabilities for AI systems that were arising from these models as these new architectures were being developed and scaled up and put to task across a really broad variety of contexts.

[00:19:25] And so as medical AI researchers, I think our first question was to start just like we did in the era of CNNs and that that first wave of discovery when Vivek and I originally started working together, there were very similar questions arising here, which is I was always taught, you know, make the care of the patient your first concern.

[00:19:44] And so the immediate questions that come to mind for me around this kind of powerful technology was in the much more challenging setting of healthcare where if a language model makes a mistake or makes an error, there's a sort of much more [00:20:00] perceptible risk or harm than in some other context, for example, in creative applications and other things.

[00:20:06] So one of the scientific questions I think, that arises in that moment when the technology is coming to fruition is, To start asking the extent to which clinical knowledge and medically important information is actually encoded in these systems to begin with. And to start asking scientific questions around how

to best measure that, but also how to begin to put metrics around it and maybe even then optimize and develop it.

[00:20:35] So at the outset of a research field, we try to make contributions that are generally useful and thoughtful, aligned with the values of the practice of medicine and of what matters to patients and people. And so that was kind of the inspiration for the first paper. And. I think the first thing we did was to look at question answering in the broadest sense, because it seemed to be a very foundational property of these, uh, Large Language Models in all of their kind [00:21:00] of foundational work.

[00:21:01] And in healthcare we took a fairly pragmatic approach. I mean, we were very lucky that this is a space and healthcare, na, natural language processing is a space in which there's been actually some fantastic work that precedes these Large Language Models. Uh, and it's great to be on the call with the likes of Andy who has been thought leading exactly that for many years.

[00:21:22] And, and I think we therefore were, were very fortunate because there are plenty of open data sets which pose medical questions and associate them with answers. As a sort of foundation for which you can start testing the ability of these models to, to do various things. So one of those capabilities that's interesting is the ability of these models to retrieve the relevant knowledge correctly.

[00:21:44] Another is the ability to manipulate that knowledge appropriately and in making an inference. And then another is the ability to communicate its conclusions in ways that are appropriate and useful and helpful to people. And so to do that we try to seek a [00:22:00] variety of data sets that, some of which encapsulate what you can think of as open domain question answering.

[00:22:05] So this is where there's a question, but then in order to answer that question, one could theoretically draw knowledge. That's not tied to a particular source. There are other conditions in which in healthcare, you might want to do closed domain question answering. For example, imagine if you have a medical research paper and you would like to have a question answered specifically about that paper.

[00:22:27] There's also then different types of knowledge in medicine. So you can imagine some settings, you may want to be answering questions about medical research in other settings like the osm, l e, you might want to be, uh, asking the kinds of questions that a healthcare professional would be asking. And then there are other settings where consumers have questions and

information is needed in lay language that's understandable about, you know, very common conditions or symptoms, for example.

[00:22:52] And so to capture that breadth, we felt that rather than focusing our research on any one of those settings, it would actually be sensible to [00:23:00] try and curate and contribute to an open body of such question answering data sets. So as I say, you know, we were very fortunate and we, in doing literature reviews, we met Dina Demner-Fushman, who's a professor at the US National Library of Medicine, and Dina and her team had curated many of these data sets and had even run public machine learning workshops and challenges to try and make progress on these.

[00:23:24] And so those included data sets, like medical question answering from consumer questions that were to National Library of Medicine. There are other data sets like PubMedQA, which provide a research abstract and you have to the question yes, no maybe. And in total there were seven of these data sets. And the seventh one, which was one we added ourselves, which we felt was also important, was of course billions of people go to the internet with their questions about their own health every day.

[00:23:51] And on Google, of course, as with many other search engines, if you put in the name of a disease or symptom, it will readily show you just externally [00:24:00] common questions that are asked about that disease or symptom. And so we were able to just using publicly available, freely available information for common diseases and symptoms obtained those questions that people commonly asked but are shown publicly on Google already.

[00:24:16] And we thought that's actually a pretty good way of starting to curate a data set that's representative of questions that are commonly that matter, that matter to billions of of consumers around the world. And so that was how we sort of set the paper up. And then the second part of the paper is not so much about the tasks, but maybe about how do you begin to evaluate these things thoughtfully.

[00:24:35] And again, there we wanted to. Start to outline some metrics that don't just look for example, at pure accuracy on a multiple choice question exam. That that, that is important and that is, that is one measure of performance. But we also felt it was really important to involve people, both clinicians, but also lay people with lived experience of diseases in evaluating different aspects of these models.

[00:24:58] And we try to do so [00:25:00] systematically. So for example, Evaluating these models by having expert clinicians rate whether or not the answer that's being provided is aligned with medical and scientific consensus. Having lay people comment about the understandability or usefulness of the answer, having metrics that reflect whether important clinical information is present in the answer and the inverse, whether it's missing in the answer and so on.

[00:25:25] So I hope that's a, sorry for the long answer. I hope that's a bit of an overview of how we set things up and why. Yeah. That that was, that was great. And it really seems that a major contribution of your paper and also. What enabled your paper in this project to take off was the existence of these public benchmarks and the creation and the curation that you did in constructing a new benchmark around the queries from the, the general public while using Google, for example.

[00:25:53] And so I think that's, it's fascinating. It's a thread in the general machine learning literature. Of course, that [00:26:00] benchmarks have really accelerated a lot of progress over the past decade and we'd love to see that more in medical AI as well. I think you started touching on this on some of the methodological contributions in addition to the benchmark.

[00:26:13] So maybe I could turn to Viva and ask you about that in particular. So your paper builds off of a long line of work. I'd love for you to highlight maybe some of the methodological advances that have been recent, uh, that have made this paper possible. And also maybe you could reflect on where you see the frontier now and the most interesting line of work to extend this going forward.

[00:26:36] Yeah, sure. I've actually been reflecting on this question over the last few days, and I think it's just to think about the progress in the language model, foundation model space over the last few years, even back in 2015's recognition. Uh, if you mention language models, I would think of ngram language models, not neural language models.

[00:26:56] And using these models to generate [00:27:00] coherent text would seem science fiction at that point of time. And so I think, as Andy mentioned, five years seems like a long time back for us in the AI community. But I think what has really catalyzed this modern, large language model revolution, uh, I think it's primarily been driven by three breakthroughs over the last few years, namely the rise of the transformer architecture, the rise of decorder only models.

[00:27:26] And lastly, I believe the development of strong alignment techniques with reinforcement learning being the cherry on the cake. So yeah, diving in, I think a lot has been said about transformers over the years, but for me, I think they're probably the biggest innovation in deep planning and AI since probably the original imaginary results back in 2012.

[00:27:47] If you look at it, it's a remarkably simple yet general purpose differentiate computer that can gobble up pretty much any kind of data that we have and run super efficiently on our hardware. And when you look under [00:28:00] the hood, the model is like super expressive in the forward pass. And there's, I think a lot that has been said about the attention layers in the model, but for me it is this beautiful generalization of the message passing paradigm that we have where each node is allowed to look at other nodes in its neighborhood, see what's interesting, and then update itself.

[00:28:20] I think that is super flexible and super general and from a computational perspective it's super useful. And if you look at the other thing in the model itself, the manner in which these attention layers the layer, layer, the feet forward layers, the residual layers that have been put together. It means that the architecture is like incredibly easy to differentiate using tools that we have at our disposal, which is, you know, great in dissent and backdrop.

[00:28:41] And lastly, the architecture has so many parallel operations. It runs remarkably efficiently on our hardware accelerators like, you know, the GPU and the tpu. And maybe one could argue that if our computer architecture itself were different, then maybe a different network would've won out over the last few years.

[00:28:56] And I think this is for something Sarah Hooker and if you has made like the [00:29:00] hardware, uh, lottery hypothesis, but I think all that is moot now. And so with this architecture, with the transformer architecture, what we have seen over the years is this remarkable convergence and adoption across domains and fields.

[00:29:11] And so I think the original paper was on translation and they kind of undersold it and they, the title itself was like very meme heavy, like attention is all you, uh, need. And I think Andre Karpathy and a few others have joked that, you know, that paper has memed its way to greatness. Since then, what you've seen is that transformers have been used in language models have been used in speech, have been used in vision, have been used in robotics and in application domains like proteins, genomics, e, everywhere you're seeing like transformer backbones and architectures.

[00:29:38] Right? So that is great. And so what this has itself enabled is the focus has shifted away from domain specific feature engineering, introducing inductive biases into the model, to focusing more on the data that these models are trained on and the compute. And so that's where the scale aspect comes in and that's where the large aspect comes in.

[00:29:57] So until 2015, 2016, I think we had language models, [00:30:00] but since then with the transformer coming through, I think the focus has been on scaling them up. And so we now have Large Language Models. And maybe one other thing that I would quickly mention is the architecture itself has been remarkably resilient over the years.

[00:30:12] I mean, transformers are now like five, six years old now, and not a lot has changed under the hood. Maybe people have like flipped around where the layer norm sits, uh, maybe they've tried to rewrite their attention kernel in a way that's more efficient depending on the kind of hardware. But, I think the consensus in the community is, you know, like keep the transformer rocket as it is and do everything else around it.

[00:30:31] Like scale up the data, the scale of the compute. And I think, I think that has led to remarkable success so far. And I think that's been, I think the backbone on which the modern large language model revolution basically has been built on. And I think what we are seeing basically is the better lesson from Richard Sutton play out over the years.

[00:30:47] Right. I think, uh, transformer is a general method that leverages computation, sorry. We're gonna come back to the scale hypothesis in the bitter lesson at the end. I think we'll just put a pin in that and we're, we're gonna tug on that [00:31:00] thread and see where we get. I think that was a great summary. Vivek, I think that there are a couple specific results from this paper that are near and dear to my heart that I'd like to drill down on, if that's okay.

[00:31:10] So full disclosure, my wife is a clinician. I saw her take step one, step two, step three. And sometime when she was studying for step two in residency, I got the bright idea that we should get an AI system to be able to do this. And I was training these very small, pathetic models like LSTM's that had 128 units and I thought that that was gonna get us past step one.

[00:31:32] But there's always been such a fascinating benchmark task for medical AI for me. And I was like so excited when I saw your paper. Cause I thought it was the first legit shot at an algorithm that could do well on step one. So just for the listeners, the summary of the result that I'd like to talk about is

you gave it a publicly available set of step one style questions that are used to prep medical students to take the exam. These are multiple choice questions. They're designed to test kind of a broad knowledge [00:32:00] base for medical students. So some of them are like, this patient walks in with these symptoms. What disease do they have?

[00:32:05] What drug should you give them? Some of them are very specific, like microbiology questions. So it is a pretty broad test, and what the test taker has to do is select the right answer from a list of potential answers, sort of weighing them against each other. So one, I would just like to say how cool it was that your model got almost 70% of those questions correct.

[00:32:24] There had kind of been a very hard ceiling around 40% and 50% before your paper. Like our models that were small and not very good, were getting like 40%. There were some Stanford papers that got close to 50%, but I thought that sort of like near passing Mark was pretty far away. So that's why I was so excited when I saw your paper.

[00:32:42] So I guess like a couple things I would like to understand are, if you did any kind of error analysis on the questions, are there blind spots or knowledge gaps or just like kinds of questions that the model gets wrong? Like are there some question formats that it gets tripped up on, or sort of what are its weaknesses when it comes to answering [00:33:00] these step one style questions?

[00:33:02] And I'll throw it, I'll just throw that to Alan. Um, or, or Vivek, actually either one. Feel free to, to hop in there. So I'm gonna pick Alan. Ah, okay. Yeah, it's, it's quite interesting. So we tried to dissect the, the responses of the model in the paper and one of the ways in which we thought it was helpful to do that would be to kind of compare its performance along these 12 axes of evaluation to clinicians.

[00:33:29] And we found that the model didn't always mention all the pertinent facts to a case, and it sometimes mentioned some facts that weren't actually relevant. One of the interesting things about that from perhaps the more AI perspective is that the palm model we set up to do it was not a retrieval enhanced model.

[00:33:48] And so you might imagine that there are then some very interesting follow-on research directions that that evaluation actually suggests might be important for the future. The other thing that we sort of noticed was [00:34:00]

that the model's uncertainty seemed to be quite a good predictor of whether it was going to get the answer right or wrong.

[00:34:07] And one of my favorite parts of the paper is a section in which, uh, Vivek and, and some of the others found a way of deferring. On some questions. So if the model, you know, setting a threshold at which it's maybe better not to answer the question. And then if you then looked at performance only on that subset, of course it was much higher.

[00:34:25] And again, I think in clinical practice, a wise doctor knows when they don't know and they know when to call a friend. And that's also been a theme of our research around responsibility in AI and healthcare and as a direction I also think is really exciting for medical ai where being responsible and knowing your limits is important.

[00:34:41] And again, I think that's another interesting area in which you could turn a limitation of the technology the other way around perhaps, and, and make it a strength. Great. I guess a follow up question could be interpreted about the model, or could be interpreted about the test. And so does Med-paLM and [00:35:00] Flam-paLM's

[00:35:01] performance on this test indicate that these questions are more testing sophisticated recall or some type of medical reasoning, and is there an meaningful difference between the two?

[00:35:13] And Alan, I'll throw that back to you as the clinician who maybe, uh, has insight into these questions. Okay. Yeah, I think that's a very interesting philosophical question and one that maybe needs people who are much clever than me and, and understand theories of medical education and state of mind much more than I do.

[00:35:31] I do think in the most simple terms, you might imagine that to answer a multiple choice question like that firstly requires a pure recall of some underlying facts and then some kind of inference and logical manipulation to reach the answer. However, you know, I'm acutely aware that AI systems don't always, you know, we can't necessarily anthropomorphize them and they don't always go about tasks seemingly the same way we do, no matter how tempting it's to imagine it.

[00:35:58] And the lessons of convolution, [00:36:00] neuronets have shown that repeatedly. The second thing is that even, okay, now if we give into that temptation and start talking about how humans do it and pretend that that's.

Perhaps relevant in some way. I think it's also true that for many clinicians over the years as you start practicing, while it's tempting to imagine that every interaction is the recall of some basic science and some underlying principles, and then deriving an answer from first principles, many times I think what makes experienced clinicians more efficient and frankly more comfortable in looking after people day to day, is that there is an element of pattern recognition in pure machine learning terms.

[00:36:34] If that was the only way the answers were solved, it might suggest that there was some leakage between the training and test sets. And we were quite careful insofar as we could be to ensure that the test we were performing, the test set was of unseen material. But what of course, perhaps can't be excluded from models at this scale is that some kind of similar concepts or patterns have occurred in language data that [00:37:00] goes into, uh, the training of these models in.

[00:37:03] At some stage, and so, you know, attributing how much of solving these difficult tasks is due to retrieval versus how much of it is due to logical manipulation. I personally find slightly harder to do in this setting than in other settings like, you know, where you've seen code completion tasks or mathematical tasks and so on.

[00:37:20] Yeah, I think that's reasonable. It is a sort of an existential question, I think, in the field as to the extent that these systems are just doing very fancy, very probabilistic, very fuzzy kinds of search lookups, or if they're actually doing some type of internal symbolic manipulation and internal kind of reasoning.

[00:37:40] I don't think that we have a good handle on machine psychology yet, and maybe that's a field that we need to, to develop to better understand sort of how these machines reason and think about the world. What I would like to jump off to next is the term Foundation Model. Um, I don't know if you're a fan of this term.

[00:37:58] This is a term introduced by some [00:38:00] researchers at Stanford in a paper, sort of generally speaking, a Foundation Model is a large model that is trained in a relatively generic way that can be repurposed for downstream applications that it was not explicitly trained to do. So I guess. Is that how you think about your work in Med-PaLM, whether or not you like the term Foundation Model that is this sort of generic substrate that we can now use to solve all sorts of medical problems.

[00:38:27] And if so, uh, what's on your sort of near term time horizon to, to use this model for, and I'll throw this one to Vivek. Yeah. I do actually think it was a bit of a clever marketing term from Stanford, or maybe that's a bit too spicy for this audience. I don't know. Uh, but jokes aside, I thought that original paper itself was really nice and gave me a very good metal model to think about the space.

[00:38:46] And I will admit that I have also used that, um, opportunistically in, uh, few different contexts. And I agree with your definition, Andy. it's a little bit fuzzy, I would say, but for me it is again, uh, a large scale pre-trained model. Uh, often trained using self [00:39:00] supervised, unsupervised learning. Uh, and this model you can rapidly apply in a bunch of different downstream settings and applications using relatively little amounts of data.

[00:39:08] So even before Med-PaLM, I think if you think about PaLM itself, I would say that's a very good example of a Foundation Model that fits broadly within the definition that, uh, we both seem to agree on. So if you see over the last year that the PaLM model, uh, that has not just been used on language tasks and benchmarks like BIG-bench, but also math and science problems.

[00:39:29] So there was this paper called Minerva, uh, from a few colleagues at Google Research, uh, medicine with our work on Med-PaLM, and also robotics, uh, with, uh, another model called PaLM-SayCan where the policy model was itself derived from Palm. So I think that's a very good example of a single foundation model that has been applied in many downstream applications with, I would say relatively little amounts of data.

[00:39:50] I think both Minerva, our application in Med-PaLM, is also similar. I think that the amount of downstream task specific data that we use is fairly small. I think that's, [00:40:00] that's a good starting point. And I think, uh, I think I think about foundation models, or at least the definition that we are using as a bridge from narrow AI to general AI.

[00:40:09] So we are somewhere in between where, uh, it's not truly maybe general AI, but it is helping us get there, uh, in some ways or, and it's doing things that are, I think, broadly useful across many different applications. And so I think our goal with Med-PaLM is also kind of similar or whatever future siblings or other variants of this model that we cook up, is we want this model to be as generally and broadly applicable across a bunch of different biomedical tasks.

[00:40:37] And not just in the text domain, not just in language tasks. Because I think we all appreciate that medicine is, uh, multimodal discipline. And so we want to generalize this model to multimodal settings as well and to more natural interaction settings, uh, like make it more natural beyond even text. So that's our goal.

[00:40:56] I think we want to make this as foundational as possible. Got it. I might [00:41:00] ask just one follow up there, and I think it's really. More of a question for myself than maybe for you guys, but Foundation Models I think are clearly beyond the scope of your traditional academic lab to build and create that they require enormous amounts of computing power.

[00:41:16] And I think often an underappreciated fact is an an enormous amount of engineering expertise. I think if you read a paper from Facebook, they published the logs of what it took to actually train that model. And it essentially is like a hundred pages of misery as far as I can tell, where a node goes down, the model won't converge, you don't know why.

[00:41:34] And there's just like clearly, um, some frustration that comes across in those logs and it's a team of very highly skilled engineers. So I wonder if you could just provide a little bit of thoughts on, you know, how. Models like these fit into a traditional academic research ecosystem. One model I have that maybe you agree or disagree with is these are kind of like particle accelerators.

[00:41:56] So I kind of think that we're in like particle physics now and there are these big [00:42:00] instruments that get built once and then we use them to interrogate various questions. Is that a good mental model for how you think about private, public, private academic research collaborations? Um, yeah. I, I, I think so.

[00:42:10] Uh, while, yeah, today it feels like, you know, these models can only be built when, you know, uh, in industrial settings and industrial research labs. I, I do still think that we are very early in terms of understanding the capabilities of these models. So a few of my colleagues like to talk about this phenomenon of emergence.

[00:42:31] I think understanding AI is going to be its own discipline. So, and I think this was, I think really well put by Demis Hassabis in one of his interviews where you said AI is one of those disciplines where we are building out the system that we want to study at the same time, and so, I think we are still mostly concentrated on the building out phase, but it may soon be that, you know, we, that itself matures.

[00:42:52] That becomes, say, more of an engineering discipline and the science transfer over to like the empirical analysis and understanding the capabilities of these models and the emergence phenomenon. I [00:43:00] think that is where academic institutions, uh, and such collaborations have a lot to give and contribute.

[00:43:06] And at that point of time, that becomes more of a science. And I don't think one is lesser than the other because you can have a system, but if you don't know what to do with it, then there's no point about using it at all. And so, uh, I, I think there's, we are still very early in this and I think emergence has been something that we are seeing, right?

[00:43:20] I think like as we scale up these models, we are seeing very interesting phenomena. We are seeing like maybe reasoning emerge for math and science, task for medicine tasks. So we don't know what will happen when we are going to scale these models up even further, maybe even beyond the text into multi-model and so on and so forth.

[00:43:33] So I think it's great that we have people from diverse disciplines starting to look at this and. Uh, I think that's gonna make it all this even more exciting. I think we are going to get a very comprehensive view of the capabilities of these models and maybe that we run into one, maybe some sort of a dead end over here, where maybe beyond a point of time these models are not improving.

[00:43:52] And then maybe we'll have to go back to a drawing and think about how to, like, rebuild AI in a way makes it more general. But I don't think we are there anywhere yet. There's still [00:44:00] a lot to be done. Awesome. Thanks. Yeah, thanks. I wanna ask you, uh, both a little more about Foundation Models and education.

[00:44:10] Uh, so I have two young daughters and we're teaching them about numbers and arithmetic. Despite the existence of calculators, they're also learning to read and write despite the existence of ChatGPT. Mathematica can differentiate and solve integrals analytically. It's been able to do this for quite some time, but we are still teaching high schoolers calculus, uh, and I suspect we would all agree that these are good things to still be doing and still be teaching.

[00:44:35] Andy and I have had very spirited debates during our postdoc days about statistics versus calculus and medical education, but nonetheless, I think we think these are all foundational concepts and ways of looking at the world

and asking questions that are important to be able to learn how to do. But where do the Foundation Models and especially large language models like ChatGPT and others, start to challenge some [00:45:00] aspects of education and the way we approach let's say medical education in particular. So Alan, you've gone through medical education, the traditional route. You're now an expert in medical AI, so I'd be very curious for your perspective, you know, have these developments changed your view of medical education, either during medical school or in a kind of continuing medical education capacity afterwards?

[00:45:25] Yeah. I think one of the many things that makes medicine magical is that it is itself continually evolving and changing. You know, I still remember, I still remember at Adam Brooks, which is the Cambridge University Hospitals kind of peering through after doing an operation, say like a, imagine I'd just done my first appendectomy or whatever.

[00:45:47] There's a book in every operating theater, or as you guys call it, operating room, where the details of the operation are sort of recorded in this handwritten book. Probably, that's all of course now in, in the EMR, but there are still [00:46:00] these handwritten books in the UK and I used to love, like, just like leafing through the old pages of the book just to see what had transpired in that operating theater before in the week before, in the month before.

[00:46:12] And you used to be able to see in, in the years before these operations where people had kind of essentially plucked the vagus nerve off the stomach of a patient. And within a few years of this happening in these logbooks, suddenly this operation disappeared. And of course that's because we'd realized that, you know, the offending problem, the cause of all this was basically a bacteria and therefore kind of doing these kind of very intricate operations to pluck nerves off stomachs, which had their own problems suddenly became not required.

[00:46:40] And an entire part of the higher surgical training curriculum and of the practice of surgery, I mean, things that used to be a, a large chunk of people's career expert careers suddenly disappeared. And of course was replaced by other, you know, amazing and important things. And there's always gonna be an element of the education of medicine that [00:47:00] is about keeping up with the state of art and what's the best possible thing we can do for the care of patients.

[00:47:06] Like, like that's, that's critical and important. And so medical education itself is of course, continually changing. I think society is also

continually changing and you, you know, if you look at the society we live in today, thankfully there's a lot more principles of participation and inclusion and, you know, our whole view of what even constitutes bias as revolutionized, I think in, in many societies in the last 10 years alone.

[00:47:30] So, and that itself, you see now, thankfully, changing the way that medicine, which also reflects society is, is happening. So an immediate and obvious thing is that, The technology itself, you're starting to see AI tools receive regulatory approval. That's of course, different to knowing and understanding whether the approved device actually improves outcomes when it's embedded in a workflow.

[00:47:50] And it's a little bit early, I would still say, in the uptake of medical AI as a tool in clinical workflows to know, but as a minimum, because the technology [00:48:00] is starting to be around. I think one element of medical education is that it's important to then understand a little bit how this technology works, what are its limitations, what are the training objectives of these devices.

[00:48:13] But I think just to be a little bit literate in the nature of the tool, and this I think is something that's most doctors are very comfortable with. New technology, again, is not unusual in medicine. You know, apparently physicians were agast when thermometers came around. Gradually, of course, physicians who used thermometers maybe found that they were perhaps slightly better than those who didn't, and so on and so on, and, and not all technology and not all tools.

[00:48:37] Are actually appropriate in every setting. And that's a a learning thing that the profession itself with patients is going to be discovering and optimizing over the coming decade. So I think learning the limitations, learning the principles of AI will be an important part of medicine. I think the other thing is thinking of these technologies themselves as a catalyst themselves as a tool for discovery, as a tool [00:49:00] for making medical education enjoyable.

[00:49:03] I think one of the most amazing things to me about the med paper was actually myself in prototyping some of these evaluation metrics myself, kind of interacting with the model to. To see how to evaluate its answers. I'm not afraid to admit that, you know, my medical knowledge is certainly not comprehensive.

[00:49:21] So I've taken specialist training and vascular surgery. But of course these questions we are putting into the model, were in all kinds of clinical specialties that, you know, my knowledge is nowhere near what a specialist

colleagues of mine might be. And I actually found, therefore it to be quite fun and quite educational, scrutinizing whether the model's answers were correct or not, and so on.

[00:49:39] And it felt to me a much more. Interactive and joyful experience than necessarily just looking up the answer in a book. It was more akin to chatting with a fellow student who also made mistakes and then together looking up the right answer. So I can see a really broad array of ways in which AI is going to kind of impact medical education.

[00:49:57] But one of the, just sort of two basic things are, [00:50:00] number one, I think as tools become available in the clinical workflow, it's gonna be important that we're all educated in their limitations, how to best use them, the evidence base that surrounds them. And I think the other thing which is maybe more creative is I suspect the tools themselves will find educational purposes.

[00:50:17] That's great. I just have to ask a quick follow on before we move towards some concluding questions. What's your take on, uh, Large Language Models, Foundation Models being, uh, authors or co-authors of medical papers? Are they co-authors or are they acknowledged? And we'll let Alan do this one. I we're gonna go to Vivek for a question next.

[00:50:38] Okay. Um. Give you a nice, nice easy one there, Alan. Yeah, I'm, I'm perhaps a little confused about how a model might be an author per se by ICJME criteria and so on. And I, and I have noted some of the respected journals have made statements, uh, around this. And so, yeah, for me personally, I haven't yet come across a [00:51:00] situation in which a model has met the, I think it's ICJME or icm, j e i, I never get the acronym correct.

[00:51:07] I've never come across a situation in which an AI system has fulfilled those criteria, so I, I personally am slightly confused about how it could be proposed. To be fair, until recently I hadn't come across a model that could get 70% on the USMLE. So I think that it's always important to think about how quickly things can change and in principle, if it's possible.

[00:51:28] I did just wanna say that I, I really loved your sentiment of the point of education being to sort of induce and maintain neuroplasticity. So like, the point of education is not to store facts, but actually to learn how to learn in the first place. So I think that that, that, that is certainly a timeless point about education.

[00:51:46] So, Vivek as promised, we're gonna come back and revisit the Scale Hypothesis. So I'm gonna try and succinctly state it and then try and get you on the record, um, as either in favor or against the Scale Hypothesis. So the Scale Hypothesis [00:52:00] goes something like this over the last 10 odd years. Machine learning and deep learning, and therefore, AI have principally been driven by engineering, by making the models bigger, training it on more data.

[00:52:13] There have been important breakthroughs, as you mentioned in the transformer architecture and things like that. Arguably, the transformer might be an engineering breakthrough because it's a paralyzable model, but by far what people have been doing is making models bigger and bigger, trained on more and more data using faster and faster computers.

[00:52:30] So the Scale Hypothesis is thus, if we keep doing that, then we will be able to conquer all areas of human intellectual endeavor. Really, we've distilled the problem to an engineering problem and we just need to throw more engineering cycles at it. The counter to that is that there are innate mechanisms and intelligence that people have that these large scale language models do not have Some of them being explicit reasoning mechanisms, understanding of causality and things like that.

[00:52:57] So I think we have seen a [00:53:00] lot of progress just by scaling models that we already have using more data. So as a leading engineer in this area, I'm curious on your thoughts as to do we need new stuff or do we just need bigger stuff? Yeah, so I, I think over the last decade or so in AI when being in this field, the one thing I have learned is to not make any predictions because it's just super hard to predict how the field evolves.

[00:53:28] A few things that I would want to say is I think the large language model, so the way it's trained, I mean, if you think about a GPT3, the, the training procedure is download a bunch of texts on the internet and then make a model, predict the next word. And it sounds simple, but I think it's deceptively simple.

[00:53:45] And so when you are doing this at scale and at internet scale and to do this really, really well, the model has to develop like, I think, you know, not just, you know, uh, syntactical knowledge, linguistic knowledge, but also understanding and reasoning [00:54:00] capabilities, world knowledge and accumulate knowledge about a bunch of different domains because the internet has, you know, chemistry and biology and physics and medicine and legal and stuff.

[00:54:07] So I think with this very simple next word, prediction objective on like internet scale data. What we have done is we've actually multitasked a bunch of different objectives and people sometimes tend to lose sight of that. It's not as simple as it seems on the outset. And so I think that's one of the beauties of this last language model where it, on the outside everything seems simple, but actually what happens under the hood is I think it's a little bit more complicated.

[00:54:30] And that's why I think we are seeing all these interesting facets and phenomena emerge as we talk to more optimally train these models, scale them up, understand better how to like, you know, train them. So I think that is one thing. And that observation aside, I think firstly we are maybe still not quite there in terms of understanding how to optimally train these models.

[00:54:47] I think the Chinchilla paper from Deep Mind showed that there is like the number of tokens that we are currently using to train some of these models. Maybe you don't need as big models as we are currently using right now. Or if you want to keep the same scale, then you have to scale up the number of text tokens that you're [00:55:00] using.

[00:55:00] So I think by understanding these scaling laws, we are going to like maybe get to better models for sure, even in the language domain itself. But then the other thing is, I mean, okay, maybe, uh, the text on the public internet kind of like comes out. The private internet still exists. We haven't still touched them.

[00:55:16] And I think what is going to happen is with these based foundation language models, People are going to build startups and, and with these startups, we are gonna have like data, fly wheels, people interacting with these systems. And so more data is going to come in, it's gonna be different kind of data, but I think that data is also going to feed into these systems.

[00:55:31] And then the second thing is multimodal. I mean, we haven't touched videos so far at all, and I think that's a huge source of understanding and improving AI systems. I think the good thing is with the transformer architecture, introducing and aggregating and assimilating all this data into the model itself is not that hard.

[00:55:46] I think the underlying architecture we have that, and with how the compute trends are evolving with like, you know, compute becoming cheaper and cheaper again, I don't think it's gonna be super expensive to train these

models and scale them up. So what I believe [00:56:00] is over the next few years, we are going to learn regardless.

[00:56:02] Uh, like I think it's not gonna be because of a lack of effort that we don't understand what happens when you, you push the scale hypothesis to a limit. Uh, for me it's hard to predict exactly what would happen. I expect. Improved capabilities. And I think as Andy you can to, I think we are already seeing superhuman capabilities in a bunch of different fields.

[00:56:22] One could maybe make an argument that ChatGPT or like you other similar systems are already kind of superhuman in many aspects just because of the kind of other wider area of tasks that they're able to solve into. But from a personal point of view, I think what I would also maybe want to see, and maybe some of this is already happening under the hood and it's kind of hidden away from users of these UI of these models where the model is like, you know, just generating text step by step.

[00:56:46] And we don't get to see at the underlying mechanisms of how it's generating this text actually. But what I would want to see is more deliberate system too, kind of like reasoning and planning. But I think we are gonna see more of that as we start, you know, teaching these models to make use of [00:57:00] tools, uh, make use of, you know, retrieve data from like the private internet or from other sources and like teach them to have like more deliberate planning behavior.

[00:57:08] And I think all that is going to happen. We seeing already, like, you know, prompting becoming like more and more sophisticated. So we are going to see, I think more of that as well. And so I think when you combine these things together, I don't know where we'll end up. I think it's hard to predict, but we'll know for sure in the next few years.

[00:57:20] Mm-hmm. Yeah, like Yogi Berra said, uh, making predictions is hard, especially about the future. Um, so I, I do think it's hard to extrapolate from where we are to where we'll be, uh, in five years. I wanted to ask a follow up sort of unrelated question, given that you described your trajectory as like a success of Massive Online Open Courses or MOOCs.

[00:57:41] I likewise am an electrical engineer who no longer uses his electrical engineering degree. I think about how a younger version of myself would fare in today's sort of job climate and just how competitive it is given how hard and how fiercely competitive the field is just to get a job. Uh, [00:58:00] given sort

of your purchase at Google, do you think there are things that we can do to identify talent that doesn't take traditional paths?

[00:58:08] It's kind of like you have to have five neuro papers already just to get an internship and if you can write five neuro papers, why do you need the internship? So have you thought about on your team at all how to sort of identify diamonds in the rough or people who are taking sort of non-traditional paths to ai?

[00:58:23] So I think over the last several years we've had a bunch of different programs to ensure that people who have maybe more of the non-traditional backgrounds are able to like come in and participate, uh, in AI research at Google. One of the programs that I think Google pioneered and I think has created, uh, many of the more prominent names in the field today, uh, is the AI Residency Program.

[00:58:50] And I think if you look at like, you know, people at like places that Open AI or even Stable Diffusion, the company behind that and a few other places, I think a lot of them have the [00:59:00] background of being part of the Google AI Residency Program. And that has also been adopted by, you know, Meta and Apple and, uh, others as well.

[00:59:07] And I think that's a very good way of ensuring that people who maybe have shown talent in, not necessarily within AI research itself, but maybe in other disciplines, uh, but have shown a keen interest to participate in ai, come in and contribute to the field and learn. And honestly, for me, like some of the best colleagues that I've had over the last few years at Google have actually come through this program and we've had some amazing collaborations over the years.

[00:59:29] And so I think that is one way for sure, like ensuring that we have more of these programs that cost a wider net and allow people without too much expectation of the number of applications that you have or what degrees that you have come in and contribute. But having said that, even uh, the number of people interested in AI itself today has grown, you know, massively.

[00:59:47] So the, the program itself. Has not been necessarily been able to, I think like, you know, scale up. And so that, I think for us, then the other question becomes how do you democratize access to the state-of-the-art resources for, you know, training and [01:00:00] deploying AI models, whether that's, you know, through frameworks like, you know, TensorFlow and JAX and PyTorch and others or, uh, open-sourcing models, uh, or like putting out papers

with enough details so that like can reproduce stuff and like, you know, uh, build on top of the research and so on and so forth.

[01:00:14] So I think as, uh, as researchers in the community, that's also a responsibility among us because. At the end of the day, the more people that we have working in this field, the better it is. And just more broadly speaking, I think we are all remarkably fortunate to be working in AI because it's one of this beautiful meta problem where, uh, if you make advances and contributions, uh, and a fundamental advances in ai, you can actually have a different, and a bunch of different applications, right?

[01:00:38] Like not just medicine or biology, but also like energy, material science, climate change, nuclear fusion. Yeah. And so I think more people deserve to have that opportunity. And I think it's up to us as, you know, educators and like people at the forefront of this field to like, you know, make sure that everyone has this opportunity.

[01:00:53] Awesome. Yeah, no, I think that's great. And I do think that the, the residency program was [01:01:00] very forward looking and remember thinking like, wow, that's a really great idea. And I think you got, especially in the first crop, just a wide range of people. You got like Goldman Sachs bankers, you got I think some humanities folks.

[01:01:12] And so it was really kind of a, a nice cut of society who are now AI experts. Um, Andy, can I ask you a question? Reverse? What, what is academia? Oh yeah. Let me turn the mic around real quick. Go for it. What is academia thinking about? Uh, how is academia thinking about this? Do you mean as far as like admissions to graduate programs or?

[01:01:30] Exactly. Um, so I was actually trying to get you to do my homework for me, um, because we are in the exact same problem where we have way more qualified graduate students than we can possibly admit, and we look for, or at least, you know, I can't speak for what every committee member does. I look for motivation and for potential.

[01:01:51] And potential can be demonstrated or it can be still kind of latent. But I, I'm looking for like, why you want to do this. And that given [01:02:00] access to the opportunities, you're gonna be successful. So I try not to explicitly select for just the fanciest CV. I like really wanna know that these problems are near and dear to your heart, and that you're gonna be motivated to push the edge of the knowledge forward.

[01:02:16] So, um, Raj, actually, uh, do you wanna say a few words about sort of how you think about this? Yeah, I agree. I think Andy summed up pretty well. The way I view this as well, I think what Andy said is totally true, which is that there are way more qualified students applying to be graduate students than we have slots.

[01:02:35] And I think the same is probably true at the assistant professor level, you know, sitting on search committees and at the undergraduate level. And this is a big thing. And so I think related to what Andy mentioned about sort of motivation and potential. The ability to articulate, I think a vision that is aligned with what the training program or the department is that you're applying to is a really [01:03:00] undervalued and incredibly important determinant of success.

[01:03:04] And I think it's pretty hard to do this to be honest, because you don't exactly know everything about every department that you are applying to. Right? You have a sense of it, but I think there's sort of being a fit between what the mission is of the graduate school program or of the department when recruiting faculty and you being a nice compliment to that research agenda, to the types of students that the graduate program likes to bring in is a really key determinant.

[01:03:32] And so it's gonna be different. It's gonna vary from program to program, but. This is a big challenge. Maybe I'll just pause there. So Alan, I, I think you touched upon this question, uh, earlier when we asked about Foundation Models in education. But, um, I'm hoping that you can just give us maybe some concluding advice to the early career clinicians.

[01:03:53] So the med students, the residents, the fellows in the audience, what should they know about AI [01:04:00] to help them prepare for a career in medicine? There's just the general thing that I was always taught by, you know, my mentors in, in medicine, and that's just been a principle for me for life, which is make the care of the patient your first concern.

[01:04:15] And that's kind of the, that applies to absolutely everything. I've failed to find a situation in which that doesn't tell me the right thing to do in medicine. And so if we apply that principle to your question, which is, you know, what do they need to learn about ai. I'd frame it entirely around what do you need to know about AI to ensure that the patients who you're caring for are gonna get the best possible care?

[01:04:37] So to me, on the one end of the spectrum, I truly believe it's the kind of technology that could theoretically bring about the most amazing

improvements in access to care, in the availability of expertise around the world. And there are so many ways that, as you know, young clinicians coming up in medical students, you can get involved in that, whether that's in research settings or whether that's [01:05:00] in translational and clinical settings.

[01:05:01] So on the one hand, you know, if there are opportunities to engage in the best ways to use existing tools, or alternatively to work with people like yourselves in academic integrated environments, that's amazing. And there will be all kinds of great opportunities there to shape the future. At the other end of the spectrum is I think, quite pragmatically.

[01:05:19] Like any tools that are in the hands of clinicians, it's really important to understand the principles of the tool. When it should be used, when it should not be used, what its limitations are. And some of that of course, becomes an art and becomes experiential. It becomes about the place of the tool in the workflow and a kind of range of socio-technical things.

[01:05:38] And as with everything in medicine that's about experience and like deliberate, iterative experience. It's a slightly simplistic answer, but I think the best thing to do with anything like this is do everything possible to make the patient in your care better. And AI is only useful or not useful if it actually contributes to that, frankly.

[01:05:58] Awesome. I think that's [01:06:00] the perfect note to end on there. So, uh, Vivek and Alan, I would just like to thank you both so much for joining us on AI Grand Rounds today, uh, sharing your work with us and helping us think through the implications of things like large language models on the future of medicine.

[01:06:16] So once again, thanks again, uh, from me and Raj, I just wanna say that it was a real pleasure to be on this podcast with both you, Argen and Alan as well. As we all know, I think we are entering a really special era for AI more generally and medical AI in particular. And I'm really excited just to see how everything unfolds over here and also collaborations between industry and academia and how we can shape the future to make the world better for everyone.