

**Dwarkesh Podcast #93 - AMA: Career Advice Given AGI, How I Research ft. Sholto &
Trenton**

Published - March 25, 2025

Transcribed by - thepodtranscripts.com

Dwarkesh Patel

Today, this is going to be an Ask Me Anything episode. I'm joined by my friends, Trenton Bricken and Sholto Douglas. You guys do some AI stuff, right?

Trenton Bricken

Yeah.

Sholto Douglas

We dabble.

Dwarkesh Patel

They're researchers at Anthropic. Other news; I have a book launching today, it's called The Scaling Era. I hope one of the questions ends up being why you should buy this book. But we can kill two birds with one stone. But, okay, let's just get at it. What's the first question that we gotta answer?

Trenton Bricken

So, I want to ask the flyball question that I heard before, of: why should ordinary people care about this book? Like, why should my mom buy and read the book?

Dwarkesh Patel

Yeah. First, let me tell you about the book, what it is. So, you know, these last few years, I've been interviewing AI lab CEOs, researchers, people like you guys, but also scholars from all kinds of different fields, economists, philosophers. And they've been addressing, I think, what are basically the gnarliest, most interesting, most important questions we've ever had to ask ourselves. Like, what is the fundamental nature of intelligence? What will happen when we have billions of extra workers? How do we model out the economics of that? How do we think about an intelligence that is greater than the rest of humanity combined? Is it even a coherent concept? And so, what I'm super delighted with is that with Stripe Press, we made this book where we compiled and curated the best, most insightful snippets across all these interviews. And you can read Dario addressing, why does scaling work? And then on the next page is Demis explaining DeepMind's plans for whether they're gonna go with the RL route and how much of the AlphaZero stuff will play into the next generation of LLMs. And on the next page is, of course, you guys going through the technical details of how these models work.

And then there's so many different fields that are implicated. I mean, I feel like AI is one of the most multi-disciplinary fields that one can imagine, because there's no field, no domain of human knowledge that is not relevant to understanding what a future society of different kinds of beings will look like. You're gonna have Carl Shulman talk about how the scaling hypothesis shows up in primate brain scaling from chimpanzees to humans. On the next page might be an economist trying to argue, like Tyler Cowen, explaining why he doesn't

expect explosive economic growth, and why the bottlenecks will eat all that up. Um, anyways, so that's why your mom should buy this book. It's the distillation of all these different fields of human knowledge applied to the most important questions that humanity is facing right now.

Trenton Bricken

I do like how the book is sliced up by different topics and across interviews. So it does seem like a nice way to listen to all of the interviews in one digestible way.

Dwarkesh Patel

Yeah. There's two interviews I've done that haven't been released publicly before that are in the book. So, one was that Jared Kaplan, who's one of your co-founders, and this is another example where it's like, he's a physicist and he's explaining scaling from this very mathematical perspective about data manifolds. And then on the next page you have a totally different perspective. It's like Goren talking about why did general intelligence actually evolve in the first place, what is the actual evolutionary purpose of it? And it's page by page, right? You can just get addresses. Even for me, the person who's been on the other end of these conversations, it was actually really cool to read it and just be like, "Oh, actually now I realize how these insights connect to each other."

Trenton Bricken

Yeah, the only other thing that stood out to me as well is the introduction section-

Dwarkesh Patel

The only thing that stood out to you?

Trenton Bricken

Yeah, that was really the only thing that was noteworthy. I just mean [what] stood out in accessibility is the introduction section and the diagrams for all the different inputs that enable you to train a machine learning model. Stripe Press books are also just beautiful, they have these nice side captions for explaining what parameters are, what a model is, these sorts of things.

Dwarkesh Patel

Actually, when we did our episode together, a bunch of people, I don't know if you saw this, independently made these blog posts and Anki cards and shit where they're explaining the concept because we just kind of passed over some things. And hopefully we've given a similar treatment to every single interview I've done, where you can read a very technical interview with a lab CEO or something, or an engineer, or a researcher, and then the side is like: here's like more context, here's more definitions, here's more commentary. And I, yeah, I feel like it elevated the conversations.

Sholto Douglas

So in other words, my parents will finally understand what I do for a job. They're gonna get it very well.

Dwarkesh Patel

Maybe my parents will. Because I got a book.

Trenton Bricken

All mine need to know is that my name's in a book.

Sholto Douglas

You're a co-author.

Trenton Bricken

They're like, "Cool."

Sholto Douglas

Should we get into the AMA questions?

Trenton Bricken

Let's do it.

Sholto Douglas

All right. So Brian Krav asks, "The issue you raised with Dario and occasionally tweet about relating to models not making connections across different topics, some sort of combinatorial attention challenge, what are your thoughts on that now? Do you solve it with scale, thinking models or something else?"

Dwarkesh Patel

So the issue is, one of the questions I asked Dario is, look, these models have all of human knowledge memorized and you would think if a human had this much stuff memorized, and they were moderately intelligent, they could be making all these connections between different fields. And there are examples of humans doing this, by the way. There's... Donald Swann or something like this, this guy noticed that what happens to a brain after magnesium deficiency is exactly the structure you see during a migraine. So then he's like, you take magnesium supplements and we're gonna cure a bunch of migraines. And it worked. And there's many other examples of things like this where you just notice two different connections between pieces of knowledge. Why, if these LLMs are intelligent, are they not able to use this unique advantage they have to make these kinds of discoveries? I feel a little shy, me giving answers on AI shit with you guys here. But, actually Scott Alexander addressed this question in one of his AMA threads, and he's like, "Look, humans also don't have this kind of logical omniscience", right? He used the example of, in language,

if you really thought about, why are two words connected? And it's like, I understand why "rhyme" has the same etymology as this other word. But you just don't think about it, right? There's this combinatorial explosion. I don't know if that addresses the fact that- we know humans can do this, right? The humans have in fact done this, and I don't know of a single example of LLMs ever having done it. Actually, yeah, what is your answer to this?

Sholto Douglas

I think my answer at the moment is that the sort of pre-training objective doesn't necessarily - like it imbues with this nice flexible general knowledge about the world, but doesn't necessarily imbue the skill of making novel connections or research. The kinds of things that people are trained to do through PhD programs and through the process of exploring and interacting with the world.

And so I think at a minimum you need significant RL in at least similar things to be able to approach making novel discoveries. And so I would like to see some early evidence of this as we start to build models that are interacting with the world and trying to make scientific discoveries, and modeling the behaviors that we expect of people in these positions. Because I don't actually think we've done that in a meaningful or scaled way as a field, so to speak.

Trenton Bricken

Riffing off that with respect to RL, I wonder if models currently just aren't good at knowing what memories they should be storing. Most of their training is just predicting the next word on the internet and remembering very specific facts from that. But if you were to teach me something new right now, I'm very aware of my own memory limitations, and so I would try to construct some summary that would stick. And models currently don't have the opportunity to do that. Memory scaffolding in general is just very primitive right now. I mean-

Sholto Douglas

Right, like Claude Plays Pokemon.

Trenton Bricken

Exactly, yeah, or like someone worked on it, it was awesome, it got far, but, another excited Anthropic employee then iterated on the memory scaffold and was able to very quickly improve on it. So that's one. I do also just wonder if models are idiot savants. The best analogy might be to Kim Peek. So Kim Peek, was born without a corpus callosum, if I recall correctly. Each hemisphere of his brain operated quite independently. So, he'd open a book, there'd be two pages visible, each eye would read one of the pages. And he had a perfect encyclopedic memory of everything he'd ever read. But at the same time, he had other debilitations; functioning socially, these sorts of things. And it's just kind of amazing how good LLMs are at very niche topics, but can totally fail at other ones still.

Dwarkesh Patel

I really wanna double-click on this thing of why there's this trade-off between memorization. Why does cutting it off... apparently it's connected to this debilitation, but why can't... Wiki text is like five megabytes of information. The human brain can store much more, so why does the human brain just not want us to memorize these kinds of things, and is actively pruning, and... yeah, I don't know. But we don't have to do it right now. We'll do a separate episode.

Trenton Bricken

Yeah, just one thing I'll say on that is, there is another case study of someone with a perfect memory, so they never forgot anything. But their memory was too debilitating. It'd be, like, your context window for the transformer is trillions of tokens. And then you spend all your time attending to past things, and are too trapped in the details to extract any meaningful generalizable insights from it.

Dwarkesh Patel

Yeah. Terrence Deacon, whose book you recommended, had this interesting insight about how we learn best when we're children, but we forget literally everything that happened to us when we were children, right? We have total amnesia.

And adults have this in-between where we don't remember exact details, but we can still learn in a pretty decent way. And then LLMs are on the opposite end of this gradient where they'll get the exact phrasing of Wiki text down, but they won't be able to generalize in these very, obvious ways.

Sholto Douglas

A little bit like Gwern's theory, optimizer theory, no?

Dwarkesh Patel

Yeah, I think I probably got it from that.

Trenton Bricken

Yeah, Gwern has definitely had a big influence on all this for me as well.

Dwarkesh Patel

I feel like what's underappreciated on the podcast is we have this group chat, and we also just meet up a lot in person. And all the output from the podcast just comes from you and a couple other people just feeding me ideas and nudges and whatever, and then I can just use that as an intuition pump during the conversation.

Trenton Bricken

Yeah, you're not the only one.

Dwarkesh Patel

What do you mean?

Trenton Bricken

Oh, like, I benefit immensely from just hearing what everyone else has to say. It's all regurgitation in one way or another.

Sholto Douglas

Another question?

Dwarkesh Patel

Yes.

Trenton Bricken

Maybe Rabid Monkey asks, "Imagine you have a 17-year-old brother/nephew just starting college. What would you recommend he study, given your AGI timelines?"

Dwarkesh Patel

That's so tough, right? I don't know, become a podcaster? I feel like that job's still gonna be around. It's funny, because I studied computer science, and in retrospect- at the time, you could've become a software engineer or something. Instead, you became a podcaster, it's kind of an irresponsible career move, but in retrospect, it's like... It kinda worked out. Just as these guys are getting automated.

Sholto Douglas

I get asked this question all the time, and one answer that I like to give is that you should think about the next couple of years as increasing your individual leverage by a huge factor every year.

So already software engineers will come up and say, "You know, I'm two times faster," or, "In new languages, I'm five times faster than I was last year." I expect that trend line to continue, basically, as you go from this model of, "Well, I'm working with some model that's assisting me on my computer, and it's basically a pairing session," to, "I'm managing a small team," through to, "I'm managing a division or a company". Basically, that is targeting a task. And so I think that deep technical knowledge in fields will still matter in four years. It absolutely will. Because you will be in the position of managing dozens- or, your individual management bandwidth will be maxed out by trying to manage teams of AIs.

And maybe we end up in a true singularity world where you have AIs managing AIs and this kinda stuff. But I think in a very wide part of the possibility spectrum you are managing enormous, vastly more resources than an individual could command today, and you should be able to solve so many more things with that.

Dwarkesh Patel

That's right, and I think I would emphasize that this is not just cope. Like, it genuinely is a case that these models lack the kind of long-term coherence which is absolutely necessary for making a successful company or... Just, getting a fucking office is kinda complicated, right? So you can just imagine that for sector after sector- the economy is really big, right?

Sholto Douglas

And really complex.

Dwarkesh Patel

Exactly, and so, I don't know the details, but I assume if it's a data sparse thing where you gotta know what is the context of what's happening in the sector or something, I feel like you'd be in a good position.

Maybe the other thought I have is that it's really hard to plan your career in general. And I don't know what advice that implies, because I remember being super frustrated. I was in college, and the reason I was doing the podcast was to figure out what it is I want to do. It wasn't the podcast itself. And I would go on, 80,000 Hours or whatever career advice, and in retrospect it was all mostly useless, and just try doing things. I mean, especially with AI, it's so hard to forecast what kind of transformations there will be, so try things, do things. I mean, it's such banal, vague advice, but I am quite skeptical of career advice in general.

Sholto Douglas

Well, the piece of career advice that I'm not skeptical of is put yourself close to the frontier, because you have a much better vantage point from there. Right? You can study deep technical things, whether it's computer science or biology, and get to the point where you can see what the issues are because it's actually remarkably obvious at the frontier what the problems are. It's very difficult to see...

Dwarkesh Patel

Actually, do you think there is an opportunity, because one of the things people bring up is, maybe the people who are advanced in their career and have all this tacit knowledge will be in a position to be accelerated by AI, but you guys four years ago or two years ago, when you were getting discovered or something, that kind of thing where you have a GitHub open issue and you try to solve it; is that just, that's done, and so the onboarding is much harder?

Sholto Douglas

That's still what we look for in hiring. So, you know?

Trenton Bricken

Yeah, I'm in favor of the "learn fundamentals, gain useful mental models", but it feels like everything should be done in an AI-native way, or top-down instead of your bottom-up

learning. So first of all, learn things more efficiently by using the AI models, and then just know where their capabilities are and aren't.

And I would be worried and skeptical about any subject which prioritizes rote memorization of lots of facts or information instead of ways of thinking. But if you're always using the AI tools to help you, then you'll naturally just have a good sense for the things that it is and isn't good at.

Sholto Douglas

Okay, next one. What is your strategy, method, or criteria for choosing guests?

Dwarkesh Patel

So, the most important thing is, do I wanna spend one to two weeks reading every single thing you have ever written, every single interview you ever recorded, talking to a bunch of other people about your research? Because I get asked by people who are quite influential often to be like, "Would you have me on your podcast?" and more often than not, I say no, for two reasons. One is just, okay, you're influential, it's not fundamentally that interesting as an interview prospect. I don't think about the hour that I'll spend with you. I think about the two weeks, because this is my life, right? The research is my life, and I wanna have fun while doing it. So is this gonna be an interesting two weeks to spend? Is it gonna help me with my future research or something?

And the other is, big guests don't really matter that much if you just look at what are the most popular episodes or, what in the long run helps a podcast grow. By far my most popular guest is Sarah Payne, and she, before I interviewed her, was just a scholar, who was not publicly well-known at all, and I just found her books quite interesting. So my most popular guests are Sarah Payne and then Sarah Payne, Sarah Payne, Sarah Payne because ... I have electric chairs with her.

And by the way, from a viewer-a-minute adjusted basis, I host the Sarah Payne Podcast where I occasionally talk about AI.

Trenton Bricken

That's funny.

Dwarkesh Patel

And then it's David Reich, who is a geneticist of ancient DNA. He's somewhat well-known, but he had a best-selling book, but he's not Satya Nadella or Mark Zuckerberg, who are the next people on the list. And then again, I think that pretty soon it's like you guys or Leopold or something, and then you get to the lab CEOs or something.

So, big names just don't matter that much for what I'm actually trying to do. And it's also really hard to predict who's gonna be the David Reich or a Sarah Payne, so just have fun. Talk to whoever you want to spend time researching, and it's a pretty good proxy for what will actually be popular.

Sholto Douglas

What was the specific moment, if there was one, that you realized you, that producing your podcast was a viable long-term strategy?

Dwarkesh Patel

I think when I was shopping around ad spots for a Mark Zuckerberg episode. And now when I look back on it, it's not in retrospect that mind-blowing, but at the time I'm like, "oh, I could actually hire an editor full-time, or maybe more editors than one, and from there turning into a real business". Because before people would tell me, "Oh, these other podcasts are making whatever amount of money." And I'd be like, "How?" You know?

So I have this running joke with one of my friends. I don't know if you've seen me do this, but every time I encounter a young person who's like, "What should I do with my life?" I'm like, "You gotta start a blog. You gotta be the Matt Levine of AI." You can do this. It's a totally empty niche. And I have this running joke with them where they're like, "You're like a country bumpkin who's won the lottery. And you go up to everything and everyone and just like, "Guys, a scratch pad. Get the scratch pad.""

Sholto Douglas

I do wanna press on that a bit more because your immediate answer to the 17-year-old was to start a podcast. So what niches are there? What sort of things would you be excited to see in new blogs, podcasts?

Dwarkesh Patel

I wonder if you guys think this too, but I think this "Matt Levine of AI" is a totally open niche as far as I can tell, and I apologize to those who are trying to fill it in. And so the other thing I'd really emphasize is, it is really hard to do this based on other people's advice, or to say "at least I'm trying not to fill a specific niche". If you think about any sort of successful new media thing out there, it has two things which are true: It's often not just geared towards one particular topic or interest, and two, the most important thing is that it is propelled by a single person's vision. It's not a collective or whatever. And so the thing I really want to emphasize is it can be done.

Two, you can make a lot of money at it, which is not the most important thing probably for the kind of person who would succeed at it, but still is just worth knowing that it's a viable career.

Three, that basically you're gonna feel like shit in the beginning where all your early stuff is gonna kind of suck. Maybe some of it will get appreciated. But it seems like bad advice to say still stick through it in case you actually are terrible because some people are terrible. But in case you are not, just do it, right? Like what is the three months of blogging on the side really gonna cost you? And people just don't actually seriously do the thing for long enough to actually get evidence or get the sort of RL feedback on, oh, this is how you do it; this is how you frame an argument. This is how you make a compelling thing that people will want to read or watch.

Sholto Douglas

Blogging is definitely underrated. I think like most of us have probably-

Dwarkesh Patel

So you both had blogs which were relevant. I don't know if they're actually relevant to getting-

Sholto Douglas

Not that. They were like somewhat relevant. But I think more so that we have all read almost all the blogs that do in-depth treatises on AI. Like if you write something that is high quality, it is almost invariably going to be shared around Twitter and read.

Dwarkesh Patel

Oh, this is so underappreciated. So, two pieces of evidence. I was talking to a very famous blogger you would know, and I was asking him, "How often do you discover a new undiscovered blogger?" And he was like, "Eh, happens very rarely, like maybe once a year." and then I ask him, "How long after you discover him or her does the rest of the world discover them?" And he's like, "Maybe a week."

And what that suggests is it's actually really efficient. Like... Oh, I have some more takes.

Trenton Bricken

Let's hear them. This is, this is the AMA.

Dwarkesh Patel

So I believe that slow compounding growth in media is kind of fake. Like, Leopold's situational awareness. It's not like he was building up an audience for a long time, for years or something. It was really good. Disagree or agree with it, and if it's good enough, literally everybody who matters- and I mean that literally- will read it. And it's hard to zero shot something like that. But the fundamental thing to emphasize is the compounding growth, at least for me, has been I feel like I've gotten better.

And it's not so much that somehow the three years of having 1,000 followers were somehow a compounding... I don't think it was that. I think it was just that it took a while to get better.

Sholto Douglas

Yeah, certainly when Leopold posted that, the next day, it's almost like you can picture it being stapled to the wall, so to speak, on Twitter.

Like, you know, everyone was talking about it. You went to any event for the following week, every single person in the entire city was talking about that essay. It was like Renaissance Florence or whatever.

Dwarkesh Patel

That's right. Yeah. The world is small.

Sholto Douglas

World is small.

Trenton Bricken

What would you say is your first big success? I'm trying to think back to when I first found your podcast. I distinctly remember you had your blog post on the Annus Mirabilis.

And Jeff Bezos retweeted it, I think. I'm trying to remember if it was before that or not, but, yeah. I'm curious, your answer.

Dwarkesh Patel

I feel like that was it. And it wasn't something where it was some big insight that deserved to blow up like that. It was just taking some shots on goal. They were all insight porn-y, and then one of them I guess caught the right guy's attention and, yeah. But I think that was it.

Sholto Douglas

Yeah, that's something else which is underappreciated, which is that a piece of writing doesn't need to have a fundamentally new insight so much as give people a way to express cleanly a set of ideas that they are already aware of in a broader way. And if it's really crisp and not articulate, then even still that's very valuable.

Dwarkesh Patel

And the one thing I should emphasize, which I think is maybe the most important thing to the feedback loop. It's not the compounding growth of the audience. I don't even think it's me getting more shots on goal in terms of doing the podcast. I actually don't think you improve that much by just doing the same thing again and again. If there's no reward signal you'll keep doing whatever you were doing before.

I genuinely think the most important thing has been that the podcast is good enough that it merits me getting to meet people like you guys. Then I become friends with people like you. You guys teach me stuff. I produce more good podcasts, so hopefully slightly better. That helps me meet people in other fields. They teach me more things. With the China thing recently, I wrote this blog post about a couple stories about things that happened in China. And that alone has netted me an amazing China network in the matter of one blog post. Right? And so hopefully, if I do an episode on China, I will be better as a result. And hopefully that happens across field after field. And so just getting to meet people like you is actually the main sort of flywheel.

Sholto Douglas

Interesting. So move to San Francisco?

Dwarkesh Patel

Yes. If you're trying to do AI, yeah.

Sholto Douglas

Oh, very important question from Jacked Pajeet. How much can you bench?

Trenton Bricken

You can't lie because we both know the answer.

Dwarkesh Patel

At one point I did bench 225 for four. Now I think I'm probably 20 pounds lighter than that or something.

The reason you guys are asking me this is because I've gone lifting with both of you. And I remember Trent and I were doing pull-ups and a bench. And it'd be, like, 'bench', and he'd throw on another plate or something. And then instead of pull-ups, he'd be cranking out these muscle ups.

Trenton Bricken

It's all technique. Let's make sure.

Dwarkesh Patel

So they both bench more than me. But I'm trying my best.

Trenton Bricken

Ask again in six months.

Dwarkesh Patel

Yeah.

Sholto Douglas

What's your favorite history book? There's a wall of them behind you.

Dwarkesh Patel

Oh, obviously the Caro / LBJ biographies. The main thing I took away from those books is LBJ had this quote that he would tell his debate [students]. In his early 20s, he taught debate to these poor Mexican students in Texas. And he used to tell them, "If you do everything, you'll win." I think it's an underrated quote. So that's the main thing I took away. And you see it through his entire career, where there's a reasonable amount of effort which goes by 20/80. You do the 20 to get the 80% of the effect. And then if you go beyond that to get, "Oh, no. I'm not just gonna do 20%, I'm gonna just do the whole thing." And there's a level even beyond that, which is an unreasonable use of time. This is going to have no ultimate impact, and still try doing that.

Trenton Bricken

You've shared on Twitter, using Anki. Or even, like, a Claude integration. Do you do book clubs? Do you use GoodReads? And what are you reading right now?

Dwarkesh Patel

I don't have book clubs. [Bud the SpaceBar edition] has just genuinely been a huge uplift in my ability to learn. Mostly because- it's not even the long-term impact over years, though I think that is part of it and I do regret all the episodes I did without using Speech Marking Cards, because all the insights have just sort of faded away. The main thing is, if you're studying a complicated subject, at least for me, it's been super helpful to consolidate. So if you don't do it, you feel like a general where you're like, "I'm gonna wage a campaign against this country." And then you climb one hill. And then the next day you're at a retreat, and then you climb the same hill. There might be a more kosher analogy. yeah. And then the other question was what am I reading right now?

Oh. My friend Alvaro De Menard, author of *Fantastic Anachronism*. Can I just hold it up? Actually it's right here. I hope he's okay with me sharing this. But he made 100 copies of this translation he did of his favorite Greek poet. Cavafy. Hopefully I didn't mispronounce it. Sorry, that one has a good inscription for Guern, because that's his copy, but it's super delightful, and that's what I've been reading recently.

Trenton Bricken

Any insights from it so far?

Dwarkesh Patel

Poets will hate this framing. I feel like poetry is like TikTok, where you get this quick vibe of a certain thing, and then you swipe. And then you get the next vibe, swipe... Alvaro, I'm sorry.

Sholto Douglas

No, that's interesting. How do you go about learning new things or preparing for an episode? You mentioned the one to two-week period where you're deep diving on the person. What does that actually look like?

Dwarkesh Patel

It's very much the obvious thing: you read their books, you read the papers. If they have colleagues, you try to talk to them to better understand the field. I will also mention that all I have to do is ask them questions, and I do think it's much harder to learn a field to be a practitioner than just learn enough to ask interesting questions. But for that it's very much the obvious thing you'd expect.

Trenton Bricken

"Based Carl Sagan" asks, "What are your long-term goals and ambitions?"

Dwarkesh Patel

AGI kind of just makes the prospect of a long term harder to articulate, right? You know the Peter Thiel quote about what is your 10-year plan and why can't you do it in six months?

Like, it's especially salient, given timelines. For the foreseeable future, grow the podcast and do more episodes, maybe more writing. So we'll see what happens after 10 years or something. The world might be different enough. So basically, podcast for now.

Sholto Douglas

Something you've spoken to me about, and particularly when you were trying to hire people for the podcast was what you wanted to achieve with the podcast. In what way do you want the podcast to shape the world, so to speak? Do you have any thoughts on that or... because I remember you telling me, "I really want people to actually understand AI and how this might change their lives."

Or, "what we could be doing now to shape the world such that it ends up better."

Dwarkesh Patel

I don't know. I have contradictory views on this. On the one hand, I do know that important decisions are being made right now in AI. And I do think, riffing on what we were saying about situational awareness, if you do something really good, it has a very high probability of one-shotting the relevant person, and people are generally reasonable. You make a good argument, it'll go places. On the other hand, I just think it's very hard to know what should be done. You gotta have the very correct world model, and then you gotta know how in that world model the action you're taking is gonna have the effect you anticipate. And even in the last week, I've changed my mind on some pretty fundamental things about what I think

about the possibility of an intelligence explosion or transformative AI as a result of talking to the Epoch folks.

Basically, the TLDR is, I want the podcast to just be an epistemic tool for now because I think it's just very easy to be wrong. And so just having a background level of understanding of the relevant arguments is the highest priority.

Sholto Douglas

Makes sense.

Dwarkesh Patel

What's your sense? What should I be doing?

Trenton Bricken

I mean, I think the podcast is awesome, and a lot more people should listen to it, and there are a lot more guests I'd be excited for you to interview.

Dwarkesh Patel

Gotta give me your recs.

Trenton Bricken

So it seems like a pretty good answer for now.

Sholto Douglas

Yeah. I think making sure that, like, there is a great debate of ideas on not, not just AI, but on other fields, and everything is incredibly high leverage in value.

Dwarkesh Patel

Yeah, yeah, yeah.

Sholto Douglas

How do you groom your beard? It's majestic.

Dwarkesh Patel

I don't know what to say, just genetics. I do trim it, but-

Sholto Douglas

No beard oil?

Dwarkesh Patel

Sometimes I do beard oil.

Trenton Bricken

How often?

Dwarkesh Patel

Once every couple of days.

Sholto Douglas

That's not sometimes! That's pretty often!

Trenton Bricken

But, do you have different shampoo for your head and your beard?

Dwarkesh Patel

No.

Trenton Bricken

What kind of shampoo do you use?

Dwarkesh Patel

Anti-dandruff.

Trenton Bricken

Do you condition it?

Dwarkesh Patel

Yeah.

Trenton Bricken

How often do you shave it?

Dwarkesh Patel

Who put you up to this?

Trenton Bricken

We're giving people the answers that they want.

Sholto Douglas

Big shampoo. Big beard oil.

Trenton Bricken

Yeah, you can sell some ad slots to different shampoo companies and we can edit it.

Sholto Douglas

Maybe we sold an ad slot. Who knows?

Dwarkesh Patel

Sorry, you had this idea of merch. Do you wanna explain this T-shirt idea?

Trenton Bricken

Yeah, yeah, yeah. So people should react to this. Someone should make it happen.

Dwarkesh wants merch, but he doesn't want to admit that he wants it. Or he doesn't want to make it himself because that seems tacky.

So I really want a plain white tee with just Dwarkesh's beard in the center of it. That's it. Nothing else.

Dwarkesh Patel

But you were saying it should have a different texture than the rest of the shirt.

Trenton Bricken

Oh, so when I was really riffing off it, where maybe a limited edition set can have some of your beard hair actually sewn into the shirt.

Dwarkesh Patel

Oh, my God.

Trenton Bricken

That'd be pretty cool. I would pay. I would pay for that.

Sholto Douglas

How much?

Dwarkesh Patel

I've got, like, patches all over my beard.

Trenton Bricken

Depends on how much hair. If it's like one is in there somewhere, versus the whole thing. Like, "Do I have to dry clean it? Can I wash it on the delicate setting?" But really, I think you should get merch. If you want to grow the podcast, which apparently you do, then this is one way to do that.

Dwarkesh Patel

You think beard and beard hair in the future is necessary?

Trenton Bricken

Oh, yeah. Oh, yeah.

Sholto Douglas

Which historical figure would be best suited to run a frontier AI lab?

Dwarkesh Patel

This is definitely a question for you guys.

Trenton Bricken

Oh, no. I mean - I'm curious what your take is first. You've spoken to more of the heads of AI labs than I have.

Dwarkesh Patel

Yeah, I was gonna say LBJ. Sorry, it's a question who would be best at running an AI lab or would be best for the world or...?

Trenton Bricken

Yeah, what's - what outcome do you want?

Dwarkesh Patel

Because I imagine it seems like what the best AI lab CEO succeeds at is raising money, building up hype, setting a coherent vision. I don't know how much it matters for the CEO themselves to have good research taste or something, but it seems like their role is more as a sort of emissary to the rest of the world. And I feel like LBJ would be pretty good at this. Just getting the right concessions, making projects move along, coordinating among different groups to maybe- Oh, Robert Moses.

Again, not necessarily best for the world, but just in terms of, like, making shit happen.

Sholto Douglas

Yeah. I mean, I think best for the world is a pretty important precondition.

Dwarkesh Patel

Oh, right. Who'd be best for the world? There's a Lord Ackwood quote of, "Great people are very rarely good people." So it's hard to think of a great person in history who I feel [would] really move the ball forward and also I trust their moral judgment.

Sholto Douglas

Yeah. We're lucky in many senses with the set today, right?

Dwarkesh Patel

That's right.

Sholto Douglas

Like, the set of people today are both... they try and care a lot about the moral side as well as sort of drive the labs forward.

Dwarkesh Patel

This is also why I'm skeptical of big grand schemes like nationalization or some public-private partnership or just generally shaking up the landscape too much, because I do think we're in one of the better.. I mean, the difficulty of whether it's alignment or whether it's some kind of deployment, safety risks. That is just the nature of the universe is gonna make that some level of difficulty. But the human factors in a lot of the counterfactual universes, I feel like we don't end up with people... Like, we could even be in a universe where they don't even pay lip service, there's not an idea that anybody had that you could have an ASI takeover. I think we live in a pretty good counterfactual universe, all things considered.

Sholto Douglas

- good set of game players on board.

Dwarkesh Patel

That's right. That's right.

Sholto Douglas

How are you preparing for fast timelines?

Dwarkesh Patel

If there's fast timelines, then there will be this six-month period in which the most important decisions in human history are being made. And I feel like having an AI podcast during that time might be useful. That's basically the plan.

Trenton Bricken

Have you made any shorter term decisions, with regards to spending or health or anything else?

Dwarkesh Patel

After I interviewed Zuckerberg, my business bank balance was negative 23 cents. When the ad money hit, I immediately reinvested it in Nvidia. So, that is the... sorry, but you were asking from a sort of altruistic perspective?

Trenton Bricken

No, no, just in general, like, have you changed the way you live at all because of your AGI timelines?

Dwarkesh Patel

I never looked into getting a Roth IRA.

Sholto Douglas

He brought us Fiji water before.

Trenton Bricken

Which was in plastic bottles, so...

Sholto Douglas

Dwarkesh has changed.

Dwarkesh Patel

Well, have you guys changed your lifestyle as a result?

Sholto Douglas

Not really, no. I, I just, like, work all the time.

Dwarkesh Patel

But you would be doing that anyways, or would you not?

Sholto Douglas

Ah, I would probably be going very intensely at whatever thing I'd picked to devote myself to.

Dwarkesh Patel

Yeah, yeah. How about you?

Trenton Bricken

I canceled my 401(k) contribution, so -

Dwarkesh Patel

Oh, really?

Trenton Bricken

Yeah, yeah, that, that felt like a more serious one. It's hard for me to imagine a world in which I have all this money that's just sitting in this account and waiting until I'm 60 and things look so different then.

Dwarkesh Patel

I mean, you could be like a trillionaire with your marginal 401(k) contributions.

Trenton Bricken

I guess, but you also can't invest it in specific things. And, I don't know. I might change my mind in the future and can restart it, and I've been contributing for a few years now.

Dwarkesh Patel

On a more serious note, one thing I have been thinking about is, how could you use this money to an altruistic end? And basically, if there's somebody who's up and coming, in the field that I know, which is making content, could I use money to support them? And I'm of two minds on this. One, there are people who did this for me, and it was kind of actually responsible for me continuing to do the podcast when it just did not make sense as there were a couple hundred people listening or something. I want to shout out Anil Varanasi for doing this. And also Leopold, actually, for the foundation that he was previously running.

On the other hand, the thing about what that blogger was saying, that the good ones you actually do notice. It's hard to find a hidden talent. Maybe I'm totally wrong about this. But I'd feel like if I put up a sort of grant application, I give you money if you're trying to make a blog, I'm actually not sure about how well that would work.

Sholto Douglas

There's different things you could do, though. Like, there's "I'll give you money to move to San Francisco for two months". And sort of meet people and get more context and taste and feedback on what you're doing and it's not so much about the money or time. It's putting them in an environment where they can more rapidly grow. Like, that's something that one could do. I think you do that quite proactively in terms of you deliberately introduce people that you think will be interesting to each other and this kind of stuff, so... yeah.

Dwarkesh Patel

Yeah. No, I mean, that's very fair, and I, obviously I've benefited a ton from moving to San Francisco. And it's unlikely that I would be doing the podcast- at least on AI- to the degree I am if I wasn't here. So maybe it's a mistake to judge people based on the quality of their content as it exists now and just throw money at them- not throw money, but give them enough money to move to SF to get caught up in this intellectual milieu and then maybe do something interesting as a result.

Trenton Bricken

The thing that most readily comes to mind is the MATS program for AI research. And this seems like it's just been incredibly successful at giving people the time, the funding and the social status justification, to do AI safety relevant research with mentors.

Dwarkesh Patel

Oh, and you – you have a similar program...

Trenton Bricken

We have the Anthropic Fellows Program.

Dwarkesh Patel

That's right, yeah. And I know you're probably selecting for a slightly different thing, but I assume it's gonna be power law dominated. And have you noticed a pattern among the, whether it's the MATS fellows or your fellows, who is just like, "This made the whole thing worth it"?

What's your first take on something?

Trenton Bricken

I mean, there have been multiple people who Anthropic and other labs have hired out of this program. So, I think the return on investment for it has been massive. And yeah, apparently the fellows, I think there are 20 of them, are really good.

Dwarkesh Patel

But what is the trick to making it work well or finding that one person?

Trenton Bricken

I think it's gotten much better with time, where the early fellows, some of them did good work and got good jobs. And so now later fellows the quality bar has just risen and risen and risen. And there are even better mentors now than before. So it's this really cool flywheel effect. But originally it was just people who didn't have the funding or time to make a name for themselves or do ambitious work. So it was kind of like giving them that niche to do it. Seems really key.

Sholto Douglas

You can do other things that don't have to be money. You could put out ideas for things you'd be really interested in reading or promoting.

Dwarkesh Patel

Yeah, yeah, yeah. There's something coming there.

Sholto Douglas

Okay, there we go.

Dwarkesh Patel

So if this episode hopefully will launch Tuesday at the same time as the book- by the way, which you can get at stripe.press/scaling. But on Wednesday, which is the day after, hopefully there's something useful for you here.

Sholto Douglas

Any other questions we wanna ask?

Dwarkesh Patel

The thing I have takes on, which I rarely get asked about, is distribution.

Trenton Bricken

Distribution of AI?

Dwarkesh Patel

No, sorry. Like, Mr Beast-style distribution, where people, I think rightly, focus on the content, and if that's not up to snuff, I think you won't succeed. But to the extent that somebody's trying to do similar things, the thing they consistently underrate is putting the time into getting distribution right. I just take random takes about... for example, the most successful thing for my podcast in terms of growth has been YouTube Shorts. It's a thing you would never have predicted beforehand. And they're responsible for basically at least half the growth of the podcast or something.

Sholto Douglas

I mean - I'd buy that. Why wouldn't you predict it? Like, I mean - like, I mean - I guess there's the contrast of - like - the long form deep content and, like, YouTube Shorts and stuff. But I definitely think they're good hooks. Good content.

Dwarkesh Patel

I have takes on how to write tweets and stuff. The main intuition being write like you're writing to a group chat. To a group chat of your friends rather than this formal whatever.

Trenton Bricken

What else comes to mind here?

Sholto Douglas

Well, maybe it's interesting the difference between TikTok and YouTube shorts.

Dwarkesh Patel

Oh, yeah. We've never cracked TikTok.

Trenton Bricken

Why not? Like, you've tried?

Dwarkesh Patel

Yeah. I mean -

Sholto Douglas

Tried? Have you done everything?

Dwarkesh Patel

No, I have not done everything.

Trenton Bricken

Have you read these poems? Maybe you're in a bubble bath with some beard shampoo on.

Sholto Douglas

Reading poems? That'd be incredible if you got any of that to go viral. You have to do that now!

Dwarkesh Patel

Manifest. Reading a poem, uncross your legs.

Trenton Bricken

Last episode it was the interpretability challenge, now it's Dwarkesh in a bubble bath.

Dwarkesh Patel

I gotta sell the book somehow, you know?

Trenton Bricken

But you literally do it like Margot Robbie -

Sholto Douglas

Yeah, exactly. Explaining the seed mechanism.

Dwarkesh Patel

Yeah, yeah, yeah, yeah, yeah.

Sholto Douglas

So what is scaling?

Dwarkesh Patel

And that's how you crack distribution.

Sholto Douglas

And that's how you crack distribution.

Trenton Bricken

Oh, but yeah, no, like when we did our episode, it launched and you were sharing interesting tidbits about how it was doing and the thumbnail you wanted to use and the title. And I think I even asked you to share more details because it seemed interesting and cool and subtle things. But it seemed like you also kind of just hated it. Like playing this game of really having to optimize all these knobs.

Dwarkesh Patel

So what I realized, I mean talent is everything, so I'm really lucky to have three to four editors who I'm just incredibly proud to work with. I don't know how to hire more of them. Like, they're just so good and self-directed. So, honestly, I don't have tips on how to correct that. I hired those guys. So one of them was a farmer in Argentina, one of them was a freshman math student in Sri Lanka, one of them was a former editor for one of Mr Beast's channels. The other is a director in Czechia who makes these AI animations that you've seen in the Notes On China, and he's working on more essays like that.

So, I don't know how to replicate that catch again.

Sholto Douglas

God, that's a pretty widely cast net, I gotta be honest. Damn.

Dwarkesh Patel

But they're all, goddamned, they're so good.

Trenton Bricken

And this was just through your challenges and just tweeting about?

Dwarkesh Patel

That's right. I had a competition to make clips for my podcast. I rounded up a couple of them this way. Yeah, it's hard, it's hard to replicate because I've tried... "tried," after.

Trenton Bricken

Why do you think this worked so well with the video editors? because you tried a similar approach with your chief of staff.

Dwarkesh Patel

Yeah. The difference is, with the video editor, I think there is this arbitrage opportunity where there are people... It is fundamentally a sort of, are you willing to work hard and obsess about getting better over time? Which all of them go above and beyond on, but you

can just find people in other countries who are... and it's not even about the wages. Like, I've 10Xed their salaries or something like that. It's just about getting somebody who is really data-oriented, and there is this global arbitrage there.

Whereas, with the general manager... By the way, the person I ended up hiring, and who I'm super excited to work with, is your childhood best friend. Max Farrens.

Trenton Bricken

Max is so great.

Dwarkesh Patel

He would have plenty of other opportunities. There's not this weird arbitrage where you find some farmer in Argentina.

Trenton Bricken

Yeah. But yeah, it is striking that you were looking for a while, and then just kind of mentioned offhand that Max was looking for something new.

Dwarkesh Patel

This is gonna be like a total, 12-year-old-learns-about-the-world kind of question, but I genuinely don't know how big companies hire. Because I was trying to find this person for a year, and I'm really glad about the person I ended up hiring. But it was just like, if I needed to hire 100 people for a company, let alone 1,000 people, I just do not know how to find people like this at scale.

Sholto Douglas

Yeah, I mean, I think this is the number one issue that startup CEOs have. Hiring. It's just relentlessly the number one.

Dwarkesh Patel

Yeah. And the thing I was stunned with is how it didn't seem like my platform helped that much. I got close to 1,000 applications across the different rounds of publicizing it that I did. And a lot of, I think, really cool people applied. But the person that I ended up hiring was somebody who was just a reference, like a mutual friend kind of thing. And a couple of other top contenders were also this way. So it's weird. Like, the best people in the world don't want to apply, at least to things like this and you just gotta seek them out. Even if you think you have a public platform or something.

Trenton Bricken

Yeah. Yeah, I mean, the job might just be so out of distribution from anything else that people would do.

Dwarkesh Patel

That's right, yeah.

Trenton Bricken

So Aditya Ray asks, "How do you make it on Substack as a newbie writer?"

Dwarkesh Patel

I think if you're starting from scratch, there's two useful hacks. One is podcasting, because you don't need to have some super original new take. You can just interview people who do, and you can leverage their platform.

And two is writing book reviews. Again, because you have something to react to rather than having to come up with a unique worldview of your own. There's probably other things, and it's really hard to give advice in advance. Just try things. But, those I think are just, like, good, cold starts.

Sholto Douglas

The book reviews is a good suggestion. I actually use, like, Gwern's book reviews as a way to recommend books to people.

Dwarkesh Patel

By the way, this is a totally under-supplied thing. Because if anybody has book reviews. Jason Furman is this economist who has like a thousand Good Reads reviews. And I probably have visited his Good Reads on a hundred independent visits. Same with the Gwern book reviews or something, right?

So book reviews are a very under-supplied thing, if you're looking to get started making some kind of content.

Trenton Bricken

I like that.

Dwarkesh Patel

Yeah. Cool. Thank you guys so much for doing this.

Trenton Bricken

Yeah, this was fun.

Dwarkesh Patel

We'll turn the tables on you again pretty soon.

Sholto Douglas

How does it feel being in the hot seat?

Dwarkesh Patel

It's nice. Nobody ever asked me questions.

Sholto Douglas

Nobody ever asked how is Dwarkesh!

Trenton Bricken

Cool. Yeah, yeah, super excited for the book launch.

Dwarkesh Patel

Thank you.

Trenton Bricken

The website's awesome by the way.

Dwarkesh Patel

Appreciate it. stripe.press/scaling

Sholto Douglas

Yeah.

Trenton Bricken

Cool.

Dwarkesh Patel

Thanks, guys.

Sholto Douglas

See you later.

Trenton Bricken

Thanks.