What AI is and isn't. What AI can do, and what it cannot. - YouTube

https://www.youtube.com/watch?v=4vCZ53AEgA0

Transcript:

(00:48)   2 00:01:13.200 --> 00:01:14.649   3 00:01:14.940 --> 00:01:16.309   4 00:01:34.170 --> 00:01:35.130   5 00:01:40.540 --> 00:01:41.490   6 00:01:46.670 --> 00:01:47.729   7 00:01:56.500 --> 00:01:57.759   8 00:02:10.630 --> 00:02:40.620 Professor Foster Provost: Hello! Everybody! Welcome to the tonight's Nyu Stern Fubon Center fireside chat uh part of our demystifying Ai series. Our topic for this evening's chat is what is AI, and what isn't it? What can it do, and what can it not do? I'm your host, Foster Provo. Uh! And this evening I have the great pleasure of introducing Professor Marcus. Um. Gary is a Professor Emeritus at Nyu. Perhaps the perhaps the youngest and one of the clarion voices in the current debate about   9 00:02:40.630 --> 00:03:08.560 Professor Provost: what really is the state of modern Ai. Gary was founder, and Ceo of Robust AI, and of Geometric Intelligence; the latter a Machine Learning company, acquired by Uber in two thousand and sixteen. Gary is the author of five books, including I have to read them The Algebraic Mind. Kluge, The Birth of the Mind, The New York Times Bestseller Guitar Zero, and his latest book co-authored with our NYU colleague, Ornie Davis, rebooting AI building machines we can trust, which, if you're in   10 00:03:08.570 --> 00:03:38.550 Professor Provost: the audience tonight, you probably ought to read um. Gary's been studying intelligence and learning, including AI for his entire professional career. As far as I can tell, his one thousand nine hundred and ninety-three Phd Thesis was on child language, acquisition generalization and neural networks, the Ai kind, and I also wanted to point, which I only learned about in talking to Gary prior to this about Gary's recent writings on sub-stack, which again, if you're interested in this topic, you really   11 00:03:38.560 --> 00:03:45.539 Professor Provost: uh, we really want to check out if you just type his name and and and sub-stack into into Google. You'll you'll find it. Um.   12 00:03:45.660 --> 00:04:02.580 Professor Provost: So. Um Oh, um, I I I here we have a we had four hundred and sixty registrants tonight. So that's That's That's that's That's a really nice audience. Um! So let's let's jump right in. Uh, Hi, Gary, welcome. Uh It seems like you've been uh, really busy lately, from what I've been reading,   13 00:04:02.590 --> 00:04:18.750 Professor Gary Marcus: I've been writing a lot. By the way, Professor Marcus dot substack, dot com to save you the the Google Great. Um, I've been pretty busy. Um, there's there's a lot of misinformation around ai these days, and i'm actually concerned about ai creating misinformation. And   14 00:04:18.829 --> 00:04:21.369 Professor Marcus: um. I think that   15 00:04:21.440 --> 00:04:50.030 Professor  Marcus: um one needs to have a background in cognitive science to really appreciate what's going on. I I think that there's a lot of hype out there, and it's easy if you don't really understand what cognition is to get caught up in the excitement, and it's important, I think, for people to be able to distinguish between where we are making progress and where we're not making progress. And if you just sort of think of Ai as magic being. When you think that there's progress in one place, you automatically assume that there is in others.   16 00:04:50.040 --> 00:05:09.909 Professor  Marcus: Um. One of the most famous mistakes like that, I think is that Ibm: thought, Okay, we'd be jeopardy. We have this amazing piece of Ai. We'll just go Throw it at the hardest problem we can think of which would be oncology, and it'll all be great. And you have John Kelly goes on. Charlie rose and says, You know we were going to change the world. And and   17 00:05:09.920 --> 00:05:17.580 Professor  Marcus: you know, after that the stock tanks for, like sixteen quarters or twenty six quarters in a row, because they never actually deliver the goods,   18 00:05:17.590 --> 00:05:31.079 Professor  Marcus: and in some ways that was hubris. And the hubris was around thinking that because I've solved one ai problem, I can solve them all. The reality is that every piece of cognition is different, and every problem that you want to solve in business is different.   19 00:05:31.090 --> 00:05:50.330 Professor  Marcus: And so I I like to think that I have some perspective there as someone who study both how children learn about things, and how I learn this thing learns about things, and I also like to think i'm not afraid to um speak truth to power and call bullshit, and that makes me unpopular in some quarters popular in others. And maybe that's why i'm here today   20 00:05:50.340 --> 00:06:20.120 Professor Provost: on terrific. I actually I think that's part of it, although I hadn't really realized the extent when I invited you. But I have to say that you know it was. It was a delight to actually wade through up. I haven't read it all I have to say. You know the great way to your writings, and I've been a fan of your writing before which maybe we talk about a little bit later. But I thought maybe since we have these sort of questions in the title that maybe we could start off simply by just what is Ai um, and then we can. You would sort of unpack some of these things.   21 00:06:20.130 --> 00:06:31.710 Professor  Marcus: We'll start with a hard one. It's not as easy as you think right, so so ai means our artificial intelligence. And so that should make you think of natural intelligence like we humans have, or honey bees have.   22 00:06:31.720 --> 00:06:49.629 Professor Marcus: And already, once you think about it that way you realize there's a lot of things involved in intelligence. It's not one thing. If I had one definition it might be like the ability to adaptively solve new problems or something like that. But that would be stacking the deck. I mean. Really, there are lots of facets of intelligence You have people like   23 00:06:49.640 --> 00:07:18.460 Professor Marcus: uh how our gardener, Robert Sternberg, saying that intelligence has a lot of facets, and I think that's right. It's not a one-dimensional variable. If you think about the Iq. Test it kind of initially makes you think like, there's just one number this Guy's Iq is one twelve and this other person seventy-two, and whatever as if it was just one number on a scale. But um, you know, the reality is that we have different kinds of intellectual talent, so there's verbal intelligence and mathematical intelligence,   24 00:07:18.470 --> 00:07:38.649 Professor Marcus: you know. Whatever intelligence a great basketball player has, I don't have that, so you know, be able to stuff yourself in the corner, and in particular angle. While doing all this, you know it. It rapid speed. So there's many things that go into intelligence, and we know how to build some of those into machines, and sometimes we do it by copying what humans do, and sometimes we don't   25 00:07:38.660 --> 00:08:07.879 Professor Marcus: um. So a good example where I would say it's intelligence is playing chess well, and we have machines that can play just very well. Another thing that I would say is a hallmark of intelligence is having conversation, you know coherent conversation. I would say you and I are doing that, and no machine can do that right now. We have machines that kind of fake it. But you know very quickly you realize that they are faking it like they're a machine or a person, or where they are, what day? It is right, so they can take it a little while. They can fool a stupid person   26 00:08:07.890 --> 00:08:37.859 Professor Marcus: or a naive person, but they have full an expert um, and so they're not very good at whatever it is that goes into conversation. I think the biggest hallmark of human intelligence is that even a child can learn a lot of new things very quickly. So you take a eight-year-old child, and you say I bought this new game at the store you want to play, and they're like, What are the rules? Three minutes later they're playing that game, maybe not at world class levels, but they picked up the rules by verbal instruction, how to play this new card, game,   27 00:08:37.870 --> 00:08:57.650 Professor Marcus: board, game, um, or how to do something physical, some new sport, or whatever. And we don't have machines that can really do that. Machines we have right now. Most of them take a lot of training to do whatever it is. They do often measured in megabytes or gigabytes, and sometimes taking months of training. And what have you? Um,   28 00:08:57.660 --> 00:09:10.259 Professor Marcus: There's a kind of flexibility that is, I think, a hallmark, a hallmark. Maybe not the only hallmark, but a hallmark of intelligence that machines don't have. And so, when you want to know, can a machine do this thing? Now   29 00:09:10.340 --> 00:09:38.429 Professor Marcus: you have to dive into the details. You can't just say. Well, they're smart now. They can do stuff like depends what you're talking about. They're not smart enough to be domestic robots now. So Elon Musk is talking about optimists. I think most of us would agree with The demo was was uninspiring, let's say. But um, you know. Aside from the fact that his robots were sort of stiff and moving slowly, and stuff like that, they didn't have the intelligence that would make them comfortable in your home. So like   30 00:09:38.440 --> 00:10:08.400 Professor Marcus: it was a demo of a robot pouring water. Well, actually there was no water pouring an empty water pitcher into um a plant. It was a little bit uh fake, according to some people on Twitter, but I didn't mean to go in that particular spot, but, like, you know there's a demo of it tilting the water picture. But if you actually wanted your robot to um, pour water into your plants, it would have to understand what the plants are, where they are, how much water to put in stuff like that. You have to have some sophistication about the world we don't have   31 00:10:08.410 --> 00:10:15.089 Professor Marcus: it's anywhere near that. So even if we could get the motor control right, the motors of tilting the robot arm and the   32 00:10:15.100 --> 00:10:38.579 Professor Marcus: correct fashion, we don't have the technology right now to build a generally intelligent robot that can participate in the world. On the other hand,   Professor Provost: Oh, I was just gonna Yeah, please finish. I want to get into this generally intelligent versus other things that people call We were.   Professor Marcus: We were talking about narrow intelligence. You you would be the other day. Um, where you have a different turn for um,   33 00:10:38.590 --> 00:11:08.490 Professor Marcus: you know. On the other hand, we can build a machine that can play chats. We can build machines that can place advertisements pretty well, and you know a lot more about that particular domain of application than I do. Um, what it basically turns out to be right now in two thousand and twenty-two. But this may change eventually, but right now we're better at building narrow intelligence than general intelligence. So when my wife and I had children, we created 2 general intelligences   34 00:11:08.500 --> 00:11:16.650 Professor Marcus: that are now eight, nine years old, and those general intelligences can go to school or read a book, or whatever, and acquire new things.   35 00:11:16.660 --> 00:11:34.960 Professor Marcus: Um, they not necessarily world class at anything that they do, but they can pick up a lot of different things, whereas the way that Ai mostly works right now is mostly people use machine learning with a lot of data focused on some very narrow problem, and in some cases the results are spectacular. The most   36 00:11:34.970 --> 00:11:55.159 Professor Marcus: spectacular thing I would say is alpha fold, which solves, maybe not perfectly. That solves quite well the protein folding problem. I've got a string of amino acids, and I want to know what shape is this going to be? And Alpha Fold is tailored to that problem and no other problem, and it doesn't it better than any other system ever did it before. It's really impressive,   37 00:11:55.170 --> 00:12:24.949 Professor Marcus: but it's a little like calculator, like a calculator kind of does some things. It's not very general. Um. Most of the things that we know now, arguably all of them that we know now have to build our narrow intelligences, and I think we'll talk later about what is the impact for that in the commercial world? And I think that's a good question. But just to get like um, what do they call it like level setting. Um, you know the reality of where we are is that we know how to build narrow intelligence or some problems not all   38 00:12:24.960 --> 00:12:44.659 Professor Marcus: so like drivers cars have been a failure so far. Um! Medical ai is still struggling so some some natural problems we can't solve Some we can, and then we have nothing really that yet that looks like general intelligence. We may someday there's no logical argument that says that's impossible, but we haven't made much progress there   39 00:12:44.670 --> 00:12:58.399 Professor Provost: right? Because so good. So I wanted to. I wanted to clarify that. Thank you. And um, because uh, I think what we're going to spend. Now, a bit of time is actually talking about the artificial general intelligence, because I know you have a let's say on that matter, and   40 00:12:58.810 --> 00:13:12.649 Professor Provost: it seems like that. There are things that need to be said right now, you know. First thing out of me I can't help but say right and so Um. So. Um: yeah. So let's I mean,   41 00:13:12.660 --> 00:13:19.880 Professor Provost: that's definitely i'm not gonna. I'm not gonna to stick to the plan I sort of had because we're here now. And so so you're um   42 00:13:21.060 --> 00:13:30.500 Professor Provost: as far as I've been able to see you are. And I know this actually because of other things. And one of the reasons why that I invited you. Is you seen as one of the top   43 00:13:31.220 --> 00:13:53.850 Professor Provost: critics? I'm gonna let you defend yourself if that's not the right term. Critics of something in this world, whether it's critic of Ai, or critic of something about Ai, you can you? You can tell me. And so is it fair to say that you're one. I mean you. You are definitely one of the top of whatever this thing is. And is it a a critic of Ai? Is it something else?   44 00:13:53.860 --> 00:14:12.039 Professor Marcus: Yeah, it's interesting question. I don't like the title, I mean, I I get why I get it a lot. Um, i'll give you a couple of reasons why I don't like it. Um, but I mean there's some truth today, so I mean look first of all, I am probably the the loudest critic of current. Ai, I mean that's a fact.   45 00:14:12.050 --> 00:14:35.010 Professor Marcus: Um, I don't exactly like the title, though. Um, for a couple of reasons. One is I actually like Ai. I've spent basically my whole life studying since I was eight years old, but first learned how to program on a paper computer. I've been interested in Ai basically from that moment forwards. And so yes, i'm the youngest. We're probably still him the youngest emeritus. Professor Didn't want you, but i'm not eight years old, but in my early fifties   46 00:14:35.020 --> 00:15:03.989 Professor Marcus: Um! So it's a long time I've been thinking about Ai, and I really think that it has a chance to change the world for the better. But I'm also really disappointed. I don't think that it is materialized to be the thing that I was expecting. I mean you. You watch the movie two thousand and one which I watched well before in two thousand and one, and you know i'm sorry, David. I'm afraid I can't do that. That that Um Ai was actually pretty smart, and could have conversations, coherent conversations. And we're still not there. I mean, we're you know, twenty years past that   47 00:15:04.000 --> 00:15:33.950 Professor Marcus: or driverless cars, you know. I wrote it about them in two thousand and twelve. I was one of the first people to talk about trolley problems with driverless cars, because I thought it was interesting and cool. I got a New Yorker piece called the More called Moral Machines. I said. You know school bus goes out of control. So your driverless car, you know, avoid it. And at that point I was a little bit warmer on Ai, and thought we were going to make progress faster than we we have, and you know, over and over. I've been disappointed in the whole field the history of over promising. So I think that you know you can love something, but but it's It's tough right? So you know. Maybe i'm the leading. It's tough right? So you know. Maybe i'm the leading.   48 00:15:33.960 --> 00:15:49.750 Professor Provost:  That's why I wanted to get a chance to say, like, exactly What are you? What? What? What, what are you? I don't think it's critical of something a little more. I don't think Ai is impossible.   49 00:15:49.760 --> 00:16:07.989 Professor Marcus: I think that it might be a positive force on society, but right now I think it's actually had a mixed kind of net contribution to society, and something good, some bad, some that i'm really worried about in the short term. So in the long term um what I mostly critical of is really hype at the end of the day.   50 00:16:08.000 --> 00:16:37.679 Professor Marcus: Um, I may, you know, be the leading critic of of Ai height. Um, I think that if we're going to make the right policies around, what to research around, how to regulate things and so forth, we need to be clear. I, if we go around thinking that, you know, because we built jeopardy Champion, we're going to solve medicine next week. We're delusional. If we go around thinking that we're going to put these cars on the road, and if they're really safe, we don't have the right data. We might again be.   51 00:16:37.690 --> 00:16:46.059 Professor Marcus: And so I want I want to to be clear-eyed about the risks, and clear eyed about what it is that we actually have.   52 00:16:46.080 --> 00:17:15.760 Professor Marcus: Um I think we're going to have. I I made a prediction. I guess it's not out yet, but um for wired, I think next year we'll probably have our first death attributed to a chat. But Chatbot will either tell someone to commit suicide or make them fall in love with them, and then make them feel abandoned, or something like that. Um, the chat bots that we have right Now, really, Don't, understand the world in kind of lab situations they've actually counsel suicide. They're just following statistics of words by following statistics of words.   53 00:17:15.770 --> 00:17:44.270 Professor Marcus: They're not really following ideas and concepts. So um example there was it. Somebody said to a chat that um, I think i'd like to kill myself. And the system says, Well, I think that'd be a good idea. Well, why is it doing that? It's not because the system actually has a concept of suicide or things that the person is talking to is a bad person, or has had a rough day or anything like that. It's just like they have this corpus, and people think in the Corpus. People say, I think I would like to x   54 00:17:44.280 --> 00:17:56.240 Professor Marcus: a lot, and then, you know, people often support their friends. They're like, Yeah, that'd be great. So you have a lot of that in the Corpus, and the the system is just following statistics. Isn't really following concepts.   55 00:17:56.590 --> 00:17:58.250 I I was gonna say   56 00:17:58.660 --> 00:18:28.560 Professor Provost: this to me. I mean, when I hear these things right. I think this is a huge failure of like product management. I mean, you know. Okay, there's the Ai. They're great. Okay, right. But to me there's and we've talked about this a lot. There's a difference between, as I think, a lot of our audience because they were my students sort of know right? The difference between the machine learning, algorithm the machine learned model, and then the product that ends up being built with this right? And Um, Yeah, right? And you're supposed to be trying to make you,   57 00:18:28.570 --> 00:18:43.979 Professor Provost: you know. Do you know, sort of build your products? So they have certain characteristics, you know. You might have failed in your product management. But I mean, this is, this is interesting question as to whether you sort of blame the Ai or blame the, you know.   58 00:18:43.990 --> 00:18:52.199 Professor Marcus: Yeah, let let's unpack that a little bit, because I I think there's a lot to it, and the it extends further than that. So um!   59 00:18:52.410 --> 00:18:57.449 Professor Marcus: The first thing to say is, yeah, a lot of Ai has failed as as product.   60 00:18:57.460 --> 00:19:26.920 Professor Marcus: Um, not all of it. So there there are some very successful ones. I would say that advertising placement has been a tremendously successful product. Google has made billions and billions, hundreds of billions of dollars um with product that places at so driven by Ai. So there's some very successful examples. And there's some epic fails. So Watson, as oncologist, was an epic fail, and so far one hundred billion dollars has gone into so-called level. Five self-driving cars. Today is the sixth anniversary   61 00:19:26.930 --> 00:19:44.909 Professor Marcus: uh, exactly to the day of Elon Musk uh tweeting. That Wall Street Journal had just run a story in which he said that we were a year away from Tesla driving coast to coast um automatically. So you know that's a fail. So far it might happen. But you know,   62 00:19:44.920 --> 00:20:09.529 Professor Marcus: he Musk said. In two thousand and sixteen we will have it in two thousand and seventeen. So far his team has not been able to deliver that product. Now, in that case this is the product people or the expectations of the Ceo, I mean. Sometimes, you know there there's pressure from the top down to deliver same thing with Watson and the oncologist that I don't think it was going to happen, and I don't think any product team could deliver   63 00:20:09.540 --> 00:20:30.579 Professor Provost: by the way, for the audience, the uh the Ibm. There's an an excellent mit tech review article that doesn't sort of post mortem of the Md. Anderson big Watson um the fail. And so if you're interested in that kind of stuff, you should, you should read that because it's very enlightening about why things went wrong.   64 00:20:30.590 --> 00:20:59.889 Professor Marcus: Yeah, or or come spring, you should come uh listen to me talking about that. I'm i'm working on. I don't know something about that. It'd be very interesting, but anyway, um, so that was a fail. Um, in the course of I I've been researching this, and so where I talked a lot to Such Saria, who is working on Sepsis. She she's Johns Hopkins, professor, and also a Ceo of a started up called Bayesian Health, and she walked me through sort of all the reality of of going in medicine   65 00:20:59.900 --> 00:21:17.350 Professor Marcus: from an idea to a product. So you do your machine learning and it's hard. It's hard because you've got data you collect in the lab, but the data that you collect in the lab don't necessarily match reality, or you collect some data in one hospital, but the other hospital does something different from the next hospital.   66 00:21:17.360 --> 00:21:30.419 Professor Marcus: And so you have a lot of problems where your machine learning works in the immediate context in which you gathered the data. But you go out in the world. It's like that, i'm saying about no plan survives uh the contact with the enemy.   67 00:21:30.430 --> 00:21:39.830 Professor Marcus: Machine learning in many domains is like that. Um. So you know, It's particularly like that in medicine, because there's so many procedures that are done differently in so many different places   68 00:21:39.840 --> 00:21:52.009 Professor Marcus: and so forth. So you have this thing. You think it's going to be a product, and then by the time you work out all the machine learning bugs. You have something robust. Then you go ahead and meet your customers and like, Yeah, So i'm, i'm busy. I can't use your thing.   69 00:21:52.020 --> 00:22:21.229 Professor Marcus: Um. And so you know that's another way in which not just in medicine, not just in a but anywhere where you build a product like what you think is like the coolest thing ever you know, until you go out out of the building, as Steve like likes to say um and meet your customers. You don't really know what it is that they actually need. So you know sometimes that works out well like Google and advertising really nailed it. But there are many, many cases, Ai or otherwise, where people have an idea for a product. But then it's not really what the customer.   70 00:22:21.240 --> 00:22:24.610 Professor Marcus: And so you have all of that stuff going on in Ai.   71 00:22:24.650 --> 00:22:28.630 Professor Provost: By the way you mentioned I I I want to. I want to   72 00:22:29.280 --> 00:22:48.499 Professor Provost: stop for a minute because you mentioned two applications um sort of using Ai and medicine, which I know is a whole ton of different applications. But let me just think about diagnosing whatever and self-driving cars. And I have this wonder. You know sort of guy who has some knowledge of the history of ai right I could I could say that, like,   73 00:22:48.980 --> 00:23:02.170 Professor Provost: by the way, some of the very first kind of seemingly successful Ai systems, we had Ai for diagnosis in medicine. We had the mason system, in the, in the, in the in the seventies, which was   74 00:23:02.760 --> 00:23:31.759 Professor Provost: an absolutely amazing system at the time, and as far as I know it, I mean it was, it was evaluated scientifically to be um more as accurate as the top diagnosticians. And then the other uh people that they evaluated against went down, as you would expect them to right. As far as I know, it never was. Um, uh, actually, you deployed in real practice for exactly the reasons you you just said.   75 00:23:31.770 --> 00:24:01.660 Professor Provost: And now here we are. Seventy S. Eighty S. Ninety S. Two thousand and two doesn't stand, you know. I mean forty, fifty years later it's forty years later, and you know, and we had the same kind of fails. And then we have driverless cars. I mean this notion that that that I think the public get that? Oh, drivers as cars are like a new thing. No. When I was in grad school in the eighty late eighty S. In Pittsburgh. They were driverless Cars are buying Cmu and Shenley Park driving around the roads right, you know, and so you know. Yes, they're they're smaller now, and they're They are better   76 00:24:01.670 --> 00:24:18.280 Professor Provost: right, you know. But if we think about it, instead of saying, look at the progress we've had over the last five years to look at the progress we've had over the last three decades, right? All of a sudden. Things things change. Do you think there's like a some kind of a not learning from history going on.   77 00:24:18.740 --> 00:24:23.529 Professor Marcus: Yeah, I mean, there's a few different things going on. There is genuine progress. I don't mean to   78 00:24:23.540 --> 00:24:42.549 Professor Marcus: say otherwise. Um. But progress is uneven. So Kurzweil likes to talk about all the exponentials. He is that what he calls the the accelerating exponentials, or something like that? Um, you know, some things really are that way. So the amount of data we have is exponentially increased.   79 00:24:42.560 --> 00:25:02.220 Professor Marcus: Our ability to simulate a human sentence has grown exponentially our ability to recognize features grown exponentially. But some other things Haven't. So general Intelligence Hasn't really grown exponentially. And it turns out that some problems you might actually need something like it. So driving is an example of this. So one   80 00:25:02.230 --> 00:25:31.710 Professor Marcus: possible story about driving, and I don't think the final story has been written is maybe, as you exponentially increase your data, your system will get, if not exponentially better, at least get enough better to be safer than a human driver. But another story is that you have outlier cases, and this is really the name of the game and driverless cars. Right now you have weird stuff that's just not in any data set. You make your data sets bigger and bigger. You always miss something. So remember, six years ago today   81 00:25:31.720 --> 00:25:49.500 Professor Marcus: Elon posts the tweet, saying, In a year we're going to go coast to coast um, you know. La to to New York. It never happens. Why doesn't it happen? Well, it's because so many unpredicted things happen. Um, the closest he has to that right is the summon feature where you type in point A and point B.   82 00:25:49.510 --> 00:25:52.870 Professor Marcus: What point B is where you're standing. Point A is where the car is now,   83 00:25:52.880 --> 00:26:15.789 Professor Marcus: and so you should be able to summon your car from New York to La if he delivered what he promised the product of the promises. Of course he has not to somebody. A month ago, or maybe it's two months to go, tried to summon their car at a trade show for airplanes. They were on an airplane runway. They press, summoned the car, drives across the runway straight into a three and a half million dollar jet   84 00:26:15.820 --> 00:26:16.730 right   85 00:26:16.800 --> 00:26:30.069 Professor Marcus: like It's not small. It's not like it. Ran into a bag of leads. It ran into a jet. Well, a person would never do that, because through their general intelligence, knowledge of the world they would think that's a big that's expense, but probably shouldn't run into it,   86 00:26:30.080 --> 00:26:39.209 Professor Marcus: and they would reason for that, even if it not been in their kind of data set, if you will, of actual driving experience, the paradigm that we have now   87 00:26:39.590 --> 00:27:03.489 Professor Marcus: has not approached that general purpose, reasoning, and approaches everything by. Let me throw data a problem, and sometimes that works. But sometimes it doesn't work as well. The same thing would happen if we had domestic robots like you'd solve watering the plans for these kind of planters, and Then somebody uses a base that looks like my glass, and it's out of the data set, and it's probably not going to work with the current techniques.   88 00:27:03.500 --> 00:27:20.019 Professor Provost: It seems like there's two issues here with the uh, because there's lots of these examples with self-driving cars. I think it's a great area to um um, and I promise, by the way, we'll talk about some positive things to not just just negative things. But I mean to me we have we. We have two things here. One is Um:   89 00:27:21.460 --> 00:27:35.170 Professor Provost: Well, one one might say, Yeah, but you know you have a lot of cases all the time when you're you know. I mean, if you, if you, if you're always going to have to have everything perfect, we might as well go home right and so   90 00:27:35.980 --> 00:27:47.440 Professor Provost: um probably. But I think all the real domains that I I mean. You know I could think I could think of scores of different domains, right, and you know, in this outlier cases. But I think there's   91 00:27:47.540 --> 00:27:52.640 Professor Provost: there's two things that end up being um important, right one.   92 00:27:53.690 --> 00:28:05.579 Professor Provost: Well, if the cost of making errors is very low like. So you deliver the wrong. Add to somebody right, you know, if it probably is an outlier, which means it doesn't happen very often. You're okay,   93 00:28:05.680 --> 00:28:34.210 Professor Provost: right, because the overall cumulative cost of your errors is low. If you're going to kill somebody, not so much right, you know. And so there's there is the cost of the errors then, of course, integrated over all the errors you make if you have a low cost, I mean, even in online advertising. If your system sucks, you know. You're going to be, you know, losing money right? But you know, because it costs lead to sort of the app. Yeah, I mean the worst case there is, you go out of business or something like that, whereas the worst case with driverless cars is, you kill a lot of people   94 00:28:34.220 --> 00:28:41.580 Professor Provost: exactly the worst case. You go out of business. So yeah. So even the catastrophic I mean probably to the founder of the company. That's   95 00:28:41.590 --> 00:28:57.019 Professor Provost: so. So. And then The second thing is,   Professor Marcus: It'll get a hundred million dollars for another real estate. It'll be no problem. And you,   Professor Provost: yeah, yeah, failing actually, man, to get getting you more money. But um! The The second, I think, is   96 00:28:57.430 --> 00:29:11.029 Professor Provost: uh, which is, which is kind of different, which is, can you actually essentially design the product to keep it from making certain things. And so one of the very first AI systems that I worked on was for um   97 00:29:11.040 --> 00:29:25.530 Professor Provost: dispatching technicians to help fix problems in the I work for the phone company for problems in the phone company. Okay. So this is a real system. Actually, I didn't build up the first system. The first system was actually a knowledge based system developed with old school stuff that was out there. Um at working   98 00:29:25.540 --> 00:29:36.440 Professor Provost: all the time, you know, doing really good stuff with with old school sort of expert system style stuff. And then in my job. I got hired to actually build, use machine learning to try to make things better. Right.   99 00:29:37.190 --> 00:29:38.989 Professor Provost: There were things like,   100 00:29:39.080 --> 00:30:08.099 Professor Provost: you know. If there's high voltage on the line. I don't care what your system says, you know, or if it's a hospital, you know, or things like that. So there were you So So you know there were these rules that you had to graft on top of it, right, no matter what. And so this to me, is an example of like, you know. Can you actually figure out what the bad cases would be, and then solve them some other way, which I think for self-driving cars. Once again it's really hard, because if the idea is,   101 00:30:08.110 --> 00:30:14.159 Professor Provost: if, In fact, you don't. You have low confidence like a lot of Ai systems. If you low confidence, then don't do anything   102 00:30:14.180 --> 00:30:40.119 Professor Provost: right. What's it mean to Don't do anything when you're driving down the street right? I mean. So I think I guess you you you you so   Professor Marcus: I I agree with all of that. Um, you know what's actually happening in driverless cars is they're not really driving the cars. But there's interesting technology. And so um! Some of the technology is to help humans drive better. So obstacle avoidance, keeping in a lane and stuff like that,   103 00:30:40.220 --> 00:31:01.039 Professor Marcus: and you may be limit the domain of application. You don't make the system. Do everything, because you realize your systems, Don't, know how to do everything, or you limit to certain places. So there are trials that are pretty close to actual self-driving cars, except that there's usually somebody patched in by video or something like that um in Arizona, for example, there's a whole bunch of companies, I think   104 00:31:01.050 --> 00:31:30.700 Professor Marcus: that are doing the things where, like the weather is perfect. You limit the roads. You may limit. Um, which particular? So which part of town, and you maybe make sure there is no know what we call unprotected left turn, and so forth. So you restrict the domain of application in a certain sense that's legit. In certain sense it's like disappointing. So it's it's not really this magical thing. I will have my car. Come, pick me up from the other coast, you know. Um, but it might have some value in in some legal locations, and if you can restrict them   105 00:31:30.710 --> 00:31:57.689 Professor Marcus: Um, that's okay. Of course the limiting case of that we've had for years, which is like an airport Mono-railway. Those a lot of them are fully automated, and you know it's one single route, and you you shield so that no people can get in the way and and stuff like that. There. There is an approach in China that they're trying to take, which is to say, okay, let's be realistic about what we can get out of a driverless car and build whole cities around that that's actually an interesting approach. You're not going to see that in the United States, because people aren't going to, you know,   106 00:31:57.700 --> 00:32:22.609 Professor Marcus: want their environments controlled. And same way, and not to stand for it. And same way. But um, there there is that avenue. Um! And so very much does depend on what problem you're trying to solve, as you said. And what is the cost of error? And can you somehow restrict the particular thing that you're doing um? So All of those variables go into like if i'm looking at AI as a solution to a particular problem. Um, is this right? And then i'll add one another, which is   107 00:32:22.620 --> 00:32:52.600 Professor Marcus: how stable is my problem. If I train it in location X, will it really work in location Y? So one of the problems with a lot of medical stuff is like you train in a teaching hospital at Harvard, or something like that. It's one thing, and then you go um out into the suburbs where they don't have the same um, you know, Fancy resources available. and they may run things very different. They might like order all of these tests um in Cambridge, because you have to meet the standard of care of your competitor hospitals   108 00:32:52.610 --> 00:33:04.770 Professor Marcus: and they're not going to do that out um somewhere in the country. And so what that test even means when it's taken in two different places is going to be different. And so you you wind up with this problem of   109 00:33:04.780 --> 00:33:32.649 Professor Marcus: the place where I collected the data is maybe not the place where I'm actually using my system. Or you know, I've trained on roads, but I Haven't actually trained on airport runways, which is a very different circumstance. And if you don't have a general intelligence which nobody does, and you're kind of at the mercy of the data, so if you're gonna be at the mercy of the data, then that can be okay, depending on what you're doing. It's fine again for advertisement recommendation. But you have to be aware like   110 00:33:32.790 --> 00:33:36.159 Professor Marcus: how much is my problem really going to differ from day to day.   111 00:33:36.170 --> 00:34:00.719 Professor Marcus: It's not just, for, by the way, like neural networks that are popular. Now, there are other famous cases, like you probably remember long term capital in the collapse of the Russian bond market. Right? You had all these genius, Nobel Prize winners make this this company, this hedge fund, and they didn't put something in their model so that we would shift from the assumptions that they made that, you know, cost them a billion dollars and messed up the entire world economy.   112 00:34:00.730 --> 00:34:02.440 Professor Marcus: Um, And And so   113 00:34:02.530 --> 00:34:12.689 Professor Marcus: it's the problem in general, with i'm going to build a model of the world. My model is probably going to be incomplete, and the data are going to change. And it's particularly problem. If you're   114 00:34:12.699 --> 00:34:40.930 Professor Marcus: only real structure is coming from correlations essentially in those data that you don't understand richly the kind of Ai that we really we want to aspire to, but have a richer understanding how the data collected, what kind of parameters might change the model, be able to do some kind of metacognition on itself, and we'll get there. But That's not where we are right now. It's where we are right now. Is it like I have a business problem I wanted to solve. I need to understand those things.   115 00:34:40.960 --> 00:34:56.649 Professor Marcus: One of the big successes right now. I don't know how it's going to play out commercially. Um. It has its own set of issues, but is um things like Dolly where you type in text, and then you get a painting out of it, or photo, or whatever it is you ask for um. That stuff is really genuinely cool.   116 00:34:57.080 --> 00:35:02.939 Professor Marcus: I don't know about the economics we could talk about that um there, aren't. These cost of error issues like.   117 00:35:02.950 --> 00:35:28.809 Professor Marcus: you type in. I want a picture of Foster, but you know, without all the books in the background, and you know it gives me something I like, or if I don't like it, then nobody ever sees it, and so it's It's no sweat um, and they are then just a matter of Can you protect the Ip. How much are you going to charge? Are you going to pay on the copyright and stuff like that? But the technology? It's fine. You don't have to worry about the kind of problems that i'm talking about. If you want to do medical diagnosis, if you want to be driving, etcetera, then you do.   118 00:35:28.820 --> 00:35:43.850 Professor Provost: And by the way, I mean I want to point out, is that is, that this the the latest generations of Ai, which are largely based on deep learning systems are doing, I mean, as as you point out with with Dolly, you know, just you know there's even, you know   119 00:35:44.050 --> 00:35:53.650 Professor Provost: kind of uh applications that are, I mean, they're strikingly great, and they're actually commercially viable. So, for instance,   120 00:35:53.660 --> 00:36:21.659 Professor Marcus: actually, I see you have your base in the background there, actually creating AI generated music, right? Which is actually an old thing, right? And it's been around since David Cope was doing it in the early eighties at least I did a gig, although I didn't play um. But I spoke with uh with Pat Methini, who did play uh at World Science Festival when he was doing, you know. Um improvised music with um algorithmic composition. Um accompanying him live. That was probably ten.   121 00:36:21.670 --> 00:36:32.339 Professor Provost: It's the lead music to the I mean. I've I've worked with like um if you, if you're interested to go go, for in the commercial side of this you might go check out A. M. P. R. I've I've   122 00:36:32.350 --> 00:36:50.369 Professor Provost: you know you tried their system out and um created music with it, and it's just like the It's better than what we had, I think, on the beginning of our you know, of our uh, uh, uh, when you were waiting to get into the to the um to the show here, right, you know. And then I was talking to a painter uh a a, a a uh,   123 00:36:50.560 --> 00:37:00.800 Professor Provost: a famous corporate artist uh who is a friend of mine, and he was saying that he used to get a paid a lot to do like aging of people,   124 00:37:00.810 --> 00:37:24.259 Professor Provost: and so we make it make a picture of someone and what they would look like, and he says it's just completely solved now by you know, sort of by by Ai, and that was something that you would get. So you know, I want to point out that there are lots of you know applications. And again, I think these are cases as you were, as you were saying, right Where? Yeah, you can look at the result, and you don't like it. No big deal, right? I mean, You're not, you know, paying for it in advance, you know. And you know, or anything like that.   125 00:37:24.270 --> 00:37:32.270 Professor Provost: I wonder if we could talk about Ai and sentence because I know lots of people have been interested in that lately, and so maybe um um.   126 00:37:32.310 --> 00:38:01.219 Professor Provost: Maybe you can explain, because i'm sure everybody in the audience doesn't know the story. So it's. Tell the story of the Lambda system which was claimed to be sentient, and I know that you wrote a a a reply that calls it nonsense on stilts, which is quite a counterclaim. Given that, like you know, I mean the people were not. I mean, these are intelligent people, and it didn't appeal to me to just simply be foolish, right? So you know. So yeah, So what's the story? Well, that's why this story got so much lead. So um A reportedly Washington Post covered it.   127 00:38:01.230 --> 00:38:13.079 Professor Marcus: Um! And she found this person at Google. It was Google Engineer who thought that Lambda, which is kind of like Gpd. Three. Is one of these language model systems was mentioned. Now   128 00:38:13.090 --> 00:38:38.230 Professor Marcus: what I don't think that the reporter mentioned. But what maybe I found out later is he was actually this guy. Blake Lemoine, I think, was kind of looking for something like this. There's a Youtube video of him several years earlier, talking about machine rights and so forth. So I I think there was a certain amount of will to believe there. Um! But he put together transcripts of him talking to the machine. They were edited, which is a problem.   129 00:38:38.240 --> 00:38:44.369 Professor Marcus: Um, but he portrayed them as if he thought I think he was genuine,   130 00:38:44.380 --> 00:39:14.349 Professor Marcus: that this system was sentient. Now I tried to talk to him about on Twitter what exactly he meant um by extension, and and he kept turning it back on me, saying, Well, what do you think I was like? Um, But so in the literature there are multiple definitions. One of them is trivial, which is something that can sense. So my apple watch here is sentient In that sense. it's kind of a boring sense, but it's actually more sentient to lambda, because um! My apple watch has an accelerometer, so it knows when i'm wiggly   131 00:39:14.360 --> 00:39:19.979 Professor Marcus: around stuff like that. It uh, you know, knows its position in space.   132 00:39:19.990 --> 00:39:37.009 Professor Marcus: Um lambda is no way anything has no senses. Um, It just actually is really. I called it an um a version of auto complete on steroids. All it's doing is predicting what word one might say next, given a string of words that came before um, The amazing thing is that that   133 00:39:37.020 --> 00:40:00.359 Professor Marcus: it looks somewhat compelling um to a human being. Um! But the fact is, that's all it's doing. Um. Another sense of sanction is intelligent. Um! And we already talked about some of the problems in in defining that. Um. But lambda is not, you know, particularly good at any kind of problem solving anything like that. Um. So I think it's a stretch there and then. The other sense is being self aware,   134 00:40:00.370 --> 00:40:05.790 Professor Marcus: and I don't see any argument for that at all. If you just look at the mechanism. It's not   135 00:40:05.970 --> 00:40:10.159 Professor Marcus: consulting some internal model over the world. So when it people work,   136 00:40:10.170 --> 00:40:40.129 Professor Marcus: we don't. You don't know reality in um. You know you don't have a perfect understanding of external reality. I don't either, but you have a model of it, so i'm looking at you. I see some books in the background, even though they're out of focus. I'm guessing that some of those are your books because of um the way the angle that they're held on, and the way people tend to roll. Um. All of this could be wrong, like the entire thing could be a deep fake, you know. Maybe David Chalmers has made this entire thing as a way of of ridiculing Professor Marcus. None of it's real,   137 00:40:40.140 --> 00:40:41.869 Professor Marcus: so I could be wrong.   138 00:40:41.880 --> 00:41:11.570 Professor Marcus: But you know there's a fact of the matter. I don't have direct access to the fact that but I have a model in my head. What's going on right now? I have model of your psychological state. I'm looking like, is he board? If I pissed him off if I lost him, can I make him smile and make a smile twice, and so i'm thinking about you as an entity. I'm thinking I probably can manipulate you, but I probably can't manipulate the book like I might make you laugh, but I probably can't make the book go off the shell. So I have all of this kind of theory about the world.   139 00:41:11.580 --> 00:41:31.330 Professor Marcus: Lambda Doesn't have any of that. It is really just playing the auto-complete game. And because   Professor ProvostI want to just point out for everybody because you mentioned David Chalmers, and who is one of our colleagues at a Nyu who is interested in this problem as well, and actually gave a talk about it last week, I think. But um, you know, but the um the   140 00:41:32.710 --> 00:42:02.700 Professor Provost: it's kind of also boring if we just equate sentience with intelligence right? And so, as I understand it, from what David was saying. Right? The bar for sentience is considerably lower in what the you know sort of philosopher is considered to be.   Professor Marcus: Dave and I had a lot of back and forth about this on Twitter, which I mean. So let me give you the end non-committal the the thing that he really uh committed to was that if we had a system that was as smart as a mouse   141 00:42:02.710 --> 00:42:12.060 Professor Marcus: which he anticipated would happen in the next ten years, so we didn't give an argument for that. If we had that, he said it would be a serious candidate for sentience, and then   142 00:42:12.070 --> 00:42:30.099 Professor Marcus: and Neil Seft and I pushed him on that. It turns out that serious candidate doesn't mean that much. It just means like, yeah, we should talk about it. If it happens, it's not a committed argument that says that mouse level intelligence is either necessary for sentience or sufficient, and and you know um   143 00:42:30.110 --> 00:42:37.340 Professor Marcus: Dave backed away from that didn't want to commit to those, and I think rightly so. It's very hard to actually   144 00:42:37.520 --> 00:42:57.529 Professor Marcus: find ground truth here, because we don't have like a little gadget that I could hold up and deem it at you and be like you know this is my tricorder, and and you know you're sentience, and this person over there. It looks like their sentience, but they're actually an Android. I mean, we have no independent way to verify that this is already a problem. But even if we did,   145 00:42:57.540 --> 00:43:04.550 Professor Marcus: I think I would share with Dave the notion that if sentience in the notion of awareness   146 00:43:04.650 --> 00:43:30.730 Professor Marcus: is not necessarily at least too tied to intelligence. There might be some minimal bit of intelligence that you need, but it's more about like. Can you reflect and have some phenomenology? Have a feeling of of something right, a a feeling of the color red or the orgasm, or you know it's not clear that Len does anything whatsoever to do with that sense.   Professor Provost: Yeah, Yeah. By the way, I would I I I just wanted to basically say that   147 00:43:31.530 --> 00:43:45.739 Professor Provost: that these are two different things. And so it's not really just the same conversation over again. The conversation about you know. How do you? How something is intelligence? And you have a separate, because I think a lot of times when, like before, this debate came up,   148 00:43:45.750 --> 00:43:52.700 Professor Provost: you know, with the exception of prior things about Ai systems being sent to what I think about sentience. I think about things like um.   149 00:43:53.660 --> 00:43:57.379 Professor Provost: Why wouldn't we want to torture a dog   150 00:43:57.390 --> 00:44:17.359 Professor Provost: right? Because you think of the dog as a sentient being, and this engine being. Therefore, for whatever reasons we don't need to go into right now shouldn't be tortured right? And so, then you could basically be like, Well, what about a bug like well in a bug. Maybe the bug isn't right. So there's whether we could put the line somewhere. But there's gonna be some line. We're above the line, you know. And so that was. That's sort of like there.   151 00:44:17.520 --> 00:44:26.679 Professor Provost: It's: yeah again. And I don't have another better word for it. But you know these are These are beings in the world, you know, and you know they Um,   152 00:44:26.830 --> 00:44:30.639 Professor Marcus: Yeah. And I mean, there there are people talking now about like.   153 00:44:30.650 --> 00:45:00.619 Professor Marcus: How smart would a computer need to be um before like? It would be moral to turn it off, and at least like following out a straight line of what we have. Now, I don't think there's any computer that would be so smart that I would have any problems whatsoever about turning it off. Maybe if we conceptualize all the stuff in different way, you know, I don't know um, but like Lambda is not in that space, you can turn off lambda. But nothing is going to happen. If you turn off Lambda, it just means you have one less auto-complete system in the world.   154 00:45:00.630 --> 00:45:25.780 Professor Provost: So I mean you. If you want to just summarize, because I want to get to something that you said that I really liked at the end of your piece, which doesn't really have to do with the sentience. It has to do with the fact that there's a lot of other serious questions that we that we need to do that we need that we need to answer. And I thought, those are very good topics for us to talk about next. But you know, when you were saying, It's not sentience. What are the key aspects   155 00:45:25.790 --> 00:45:30.550 Professor Provost: that are that you know what's the sort of the keys to that are making? You say? Sure   156 00:45:30.740 --> 00:45:45.009 Professor Provost: it's an auto-complete, you know. It's like It's like an an amazing auto-complete system. But you know what that doesn't mean. Necessarily. That you know that it is not something in there right, you know.   157 00:45:45.020 --> 00:45:57.559 Professor Provost: By the way, I don't believe it to be sent you just to save you that you know, with   Professor Marcus: There is a line in the um essay where I quoted from Lambda, I mean I have to do it from memory. I won't get it for Beta, but Lambda says something like   158 00:45:57.570 --> 00:46:13.009 Professor Marcus: in response to a question of what do you like to do with your time? Lambda says I like to play with my friends and family and do good things, do good deeds, and I liken this to a psychopath pretending to have friends, friends and family in order to get you to like it.   159 00:46:13.020 --> 00:46:19.120 Professor Marcus: Um for them. Um, but real psychopaths actually like,   160 00:46:19.130 --> 00:46:49.080 Professor Marcus: you know, we make stuff up, but they're actually trying to do something to manipulate you, Lambda isn't even doing that. It is making up for those friends and family. It has no reference what it means by who's friends or family? It's just words. It's just predicting sequence of words. It's like there are people that they play scrabble in English who don't speak English, and they've just memorized the English scrabble dictionary, and you know, you can do that. Lambda is like memorizing the English Scrabble Dictionary, not actually   161 00:46:49.090 --> 00:47:13.799 Professor Marcus: speaking English. It may fool you, and I could watch that that non English scrabble player and be like Wow! That was a great seven letter move, but it doesn't mean that they know the the words they're playing with. And until you get past that point where there are reference to the things that you're talking about. You invoke friends and family like, who are your friends and family like. If you're not connecting that, then I just don't even see how you're in the league of things that we should talk about.   162 00:47:13.810 --> 00:47:21.660 Professor Provost: Yeah. And so that's actually I mean, I i'm. I'm sort of very much in that mode as well. And, by the way, I just want to point out, Lambda stands for   163 00:47:22.130 --> 00:47:26.779 Professor Provost: language, model for dialogue applications like   164 00:47:26.970 --> 00:47:46.309 Professor Provost: we're looking at the language model whatever happened to the dialogue application, I mean, if you have a dialogue application, you would maybe build in some memory. So you remember what you said before   Professor Marcus: I know. I know you want to go on. But just very briefly. There's a technical term now that is spread very widely, which is called language model. You just use it. Um! And   165 00:47:46.320 --> 00:47:53.189 Professor Marcus: what it means in the field is a model of sequences of words.   166 00:47:53.200 --> 00:48:23.029 Professor Marcus: What it would mean to you as a late person might be like. I understand what you're talking about, and this is not what it is. It's a trade term like deep learning means how many layers you have in a neural network. It doesn't mean conceptual depth, and people end up getting confused by these terms. They know what they mean initially, and then they get lost. They forget what they language. Model is is just the statistics of which word follows which other? Whereas, if you've ever worked on dialogue systems to maybe you have what you want to do There, is to be able to   167 00:48:23.040 --> 00:48:39.120 Professor Provost: take ideas and turn them into sentences and sentences back into ideas. And these so-called language models close quote Don't actually do that.   168 00:48:39.130 --> 00:48:57.669 Professor Provost: No, no, my point was, thank you because I think the audience doesn't necessarily know all those things, and my part was Yes, the language model is a very useful thing if you can be building a dialogue application. But the language model Isn't the dialogue application right? And so you know the the, you know sure have a really good model of what words should follow. What other words that's going to be super helpful. If you're building a dialogue application, It's not   169 00:48:58.170 --> 00:49:01.960 Professor Marcus: it actually remains to be seen whether they can be   170 00:49:01.970 --> 00:49:31.940 Professor Marcus: control. Let's say, I think of them like bowls in the china shop. So the brilliant thing is they can talk about anything and come up with a response. It's grammatically fluent. It might even sound relevant, but it's often misinformation. Um, and sometimes it's toxic, and so forth, and is very difficult to control it. So you ask one of these systems. Why is it okay to eat socks? After meditating, and instead of saying "it's not ok to eat socks after meditating", it comes up with bullshit like um. Some experts believe that it is good to eat socks   171 00:49:31.950 --> 00:49:43.719 Professor Marcus: after meditating, because such and such it sounds like an answer. It fits the statistics of language, but it's complete nonsense, and nobody yet knows how to make these things   172 00:49:43.730 --> 00:50:12.970 Professor Marcus: avoid that problem. So if you really want to use them in a dialogue system, unless your subject is like fiction, or it's the realist pros, or something like that, you actually wind up with a problem. So if you want to use one of these things, let's say as a psychiatrist, you have to be incredibly careful, because it is prone to telling people to commit suicide and so forth. And you know a real dialogue system should have more comprehension of what what is under discussion. So yeah, like three hundred startups, or whatever using Gpt. But not that Many of them   173 00:50:12.980 --> 00:50:35.049 Professor Marcus: that I know of are like commercially getting it to actually work. Some of the best applications, I think, are like copyrighting things. They're really like great for high school students who want to plagiarize basically like you type something, and it's spit something out. It might be true. It might not be. It kind of looks. Okay, Um. But if you actually want to use it, you have to like fact check the thing. You can't trust it, and   174 00:50:35.060 --> 00:51:04.260 Professor Provost: so forth. So you know it's most. I just because I wanted to get a couple of questions questions from the from the Um uh, Q&A. Which actually are sort of right along lines with what you were saying at the end of this thing. There's a lot more important things that we we want to talk about right. And so um while he talked about the the failures that involve things like labeling black people as as gorillas, you know. And then Matthew is talking about the issues about like Ai as a as a judge,   175 00:51:04.270 --> 00:51:21.739 Professor Provost: you know, or sort of to, as a helpful in the you know, in the in these systems where they're adjudicating, you know. Um uh i'm putting words in um in in Matthew's mouth. But some of these um uh bail hearings, and and whether or not you're they're going to um uh the recidive recidivism risk and things like that. And so,   176 00:51:22.150 --> 00:51:29.749 Professor Provost: when we actually making decisions right, we need to be able to trust the systems,   177 00:51:29.760 --> 00:51:51.220 Professor Provost: right and uh part of that trust, when we, when at least what I'm thinking about, do I trust a person to be able to make decisions? It's such that I would put them in a place where they're going to be Decision Making part of that is also the ethical issues right? Do I believe that they have the right ethical and moral foundations that they're probably going to make the right decisions, although people make mistakes sometime.   178 00:51:51.470 --> 00:52:06.559 Professor Marcus: Yeah, I I um. Where Are we a lot about putting these things in the legal context? Um, For multiple reasons. One is, they don't actually have the capacity to reason over moral codes, legal codes, and so forth.   179 00:52:06.570 --> 00:52:16.570 Professor Marcus: Um! And so that's inherently problematic stack, and the cost of error can be high in legal situations. You know somebody can have a death penalty or   180 00:52:16.580 --> 00:52:46.520 Professor Marcus: um, you know, falsely and so forth. Um, Then there are issues of bias. So by and large current Ai systems perpetually past history that's not always the appropriate thing to do. Sometimes past history is based on bias, discrimination, et cetera. And these systems don't really understand that and can't really get past it. So it worries me a lot. I notice. One of the questions was about Google's spectacular failure with a black people being labeled as as gorillas. Well, that was like six years ago   181 00:52:46.530 --> 00:53:13.960 Professor Marcus: that that was solved. That doesn't happen anymore. But there are so many more problems like that that just   Professor Provost: could solve or just patched just by the way?   Professor Marcus: Patched is a great word for it. so individual little problems get patched. That particular thing is like they put a bandaid on the like month ago, or somebody somebody uh tweeted as an example, something that just happened which was they Google for Ceo and all the top 50 hits were mail   182 00:53:13.970 --> 00:53:23.089 Professor Marcus: that doesn't actually reflect reality, like the Ceo of Oracle Big Company is female, but she didn't pop up in that search, and you know it's top fifty company.   183 00:53:23.100 --> 00:53:51.160 Professor Marcus: Um. So there's bias in the data set. So um in statistics we sometimes talk about convenient samples. You sort of get whatever data happens to be available, most of the Internet is a half-assed convenient sample. It's like whatever people will happen to put up there. And so you know, maybe the favorite movie had a male Ceo, and then that gets represented more than it actually is in reality, because we had a cultural bias. And so a lot of things just get exasperated.   184 00:53:51.170 --> 00:53:57.099 Professor Marcus: Um exaggerated both of those at the same time. Um! And get get worse and worse. And so   185 00:53:57.290 --> 00:54:19.520 Professor Marcus: there is no general solution so far to those problems. You just pass them up. People notice them. You fix that you pass the next to me.   Professor Provost:I wanted to get to the point of asking you about like, what are recommendations since time is unfortunately running out. And I want to say, I mean, like some of this stuff, you know, eh? First of you know some of the stuff My reaction is, you know it like,   186 00:54:20.390 --> 00:54:35.230 Professor Provost: Where did your data Scientists go to school kind of stuff like, you know. I mean, you know, we think about. You know what We teach our data scientists just to actually be really really careful about the selection of the training data that you use to actually build your machine learning models. But of course I don't think that's really.   187 00:54:35.240 --> 00:54:41.799 Professor Provost: You know the the the solution to these, you know. So what's your what? What are your recommendations?   188 00:54:42.120 --> 00:54:55.520 Professor Marcus: I mean? I I think of this line from AA, which is recognition as the first step. I think that the field is having trouble, recognizing that it's really Discoveries are not, in fact, sufficient,   189 00:54:55.530 --> 00:55:10.899 Professor Marcus: and that there's a kind of power dynamic in the field where people have these deep learning discoveries have almost all of the power, almost all the money, and are telling other people to shut up literally, and that's not good. Um,   190 00:55:11.240 --> 00:55:26.329 Professor Marcus: I think that we need to recognize that there are limitations, and we need to foster innovation, even if that means sometimes not going one hundred percent for the low hanging fruit. But it's very hard to implement that   191 00:55:26.340 --> 00:55:47.309 Professor Marcus: um, and you know what it might take is some spectacular failure. So maybe at some point investors would be like We put a hundred billion dollars into driverless cars, and you're crashing into airplanes on runways like, What is this? We're pulling out our money, and maybe that, you know, makes an impression. Or they say you built these chat bots, you know, Five years from that we might have all this gpt stuff,   192 00:55:47.320 --> 00:56:04.460 Professor Marcus: and it's like, Well, what do you actually using it for? Is it really go into product, you know It's so. It might be that we actually need to have something bad. Another Ai winter before people settle down and realize that cognition is not homogeneous. Intelligence is not one thing, but many   193 00:56:04.470 --> 00:56:34.450 Professor Marcus: that we've licked some of it, and not other parts, and we need to go hard into those other parts, and not keep playing this thing of well, check with me in five years, or check with me in three months. I'm sure I've got it like. There's been a lot of that kind of promise, right? You, Shanley Park. Um driverless car experience in the eighties like this. There's been sixty years of people saying, You know, next week, or in five years, there has to be some realization that hey? We don't actually have all the answers. I mean, Ultimately, I'm talking about humility.   194 00:56:34.740 --> 00:56:54.639 Professor Marcus: I think that right now the field lacks humility, and until that humility is front and center and graduate students are encouraged to work on hard problems rather than just little differential tweaks to what's out there. I'm not super optimistic. I'm a little bit more optimistic than it was, because   195 00:56:54.960 --> 00:57:04.110 Professor Marcus: there's starting to be some recognition that there's starting to be a little bit more oxygen to neuro symbolic Ai, which I think is part of the answer. Not all of it, but part of the answer.   196 00:57:04.120 --> 00:57:27.489 Professor Marcus: Um. And so, if we can give more oxygen to ideas that aren't as popular, I think it'll help. And, in fact, new, deep learning was in around two thousand and ten was completely out of favor. And it's because organizations like Cfar gave them funding even when they were less popular, that they were eventually able to thrive when their moment came, which I actually had to do with independent things about Gpus, allowing them   197 00:57:27.500 --> 00:57:30.819 Professor Marcus: to work faster, but we need to learn the lesson that   198 00:57:30.830 --> 00:57:57.990 Professor Marcus: of not well. Sometimes the phrases use premature closure, so we didn't prematurely close, deep learning, and eventually got up the oxygen it needed. I think a lot of other ideas are being shut down because you can. In most things you can make a good quick approximation with deep learning, and fool yourself into thinking that you're going to get all the way there. The metaphor that I I borrow from Youber Dreyfus is a ladder to the a ladder doesn't necessarily get you to the moon.   199 00:57:58.000 --> 00:58:16.190 Professor Marcus: We've built a lot of fantastic ladders, but we have to recognize that sometimes we're going to need other techniques to get to the things that we aspire to do. That means,   Professor Provostlet me let me interrupt you just because we have. We do have some. A bunch of questions came in um around a few minutes ago, so Art asks, Do you think um   200 00:58:16.200 --> 00:58:22.519 Professor Provost: efforts uh such as our Cuban to popularize and teach Ai to teenagers are contributing to the overall.   201 00:58:22.630 --> 00:58:50.210 Professor Marcus: I can't speak to Cuban's efforts. I don't know anything about them, but I will say that I worry a lot about the popular media, and I think that there's a triumph Narrative. That's very easy to sell, and gets clicks and and the media loves. But the the Triumph narrative is always the true one. Sometimes you know, the best paper in any given week is probably the one that says we tried this cool idea, and it worked a little bit, and we're going to keep working on it, and that's really dull.   202 00:58:50.220 --> 00:59:14.479 Professor Marcus: And so Then, when you know one of the Fang companies comes out with a press release and says, we've solved it. Then that's what the media popularizes. And um over time. I think that creates a lot of bias I have noticed lately. The cade mats who used to do that kind of thing is really pushed back on the driverless car stuff lately, and that's that's a good thing. When you have um A journalist who, I think, outgrows the initial infatuation and says, Hey, hang on And where is that delivery that we were promised?   203 00:59:14.490 --> 00:59:24.880 Professor Provost:  So Deborah actually about driverless cars going back to our discussion before it says, Well, I mean, what if the driverless car is one hundred and fifty,   204 00:59:25.150 --> 00:59:39.509 Professor Provost: on average, are much safer than human beings, and then once in a while they they make a stupid mistake, and you know and kill somebody. How do you balance that? How do you?   Professor Marcus: I change the word from? If to when I think it will happen, but I don't think it's happened yet.   205 00:59:39.520 --> 00:59:52.060 Professor Marcus: Um, I think right now. The driverless cars, if you put them in just arbitrary driving drive anywhere would be no match for human beings. They would have a lot of trouble with weather they'd have a lot of trouble with   Professor Provost: what if they're unlimited access highways   206 00:59:52.070 --> 01:00:21.790 Professor Marcus: where there are unlimited access highways. I recommend that you read an article that Kate Mats wrote a few weeks ago about trucks. That's That's as close as anybody has come to this, and even then it's still not really working that Well, I mean it really isn't as good as humans, of course, also with trucks. The cost of there is even bigger um. And so you really got to get it right. Um! It's very disturbing that Tesla won't release It's intervention statistics because they have the the biggest fleet, but they go around telling the State of California. We don't have to.   207 01:00:21.800 --> 01:00:35.049 Professor Marcus: give you data on driverless cars, because our cars, even though we call them full self-driving and autopilot, are not actually the autonomous vehicles. And so the the biggest source of data would be there, and they're not sharing um, so you know,   208 01:00:35.060 --> 01:01:04.999 Professor Marcus: even if it were true, the driverless cars are safer within people, which i'm sure pretty sure it's not true. We wouldn't actually have the right data to know it. And the fact that that people who have those data and Don't want to share them is is, you know, disconcerting. Let's say so. But anyway, going back to the question, there will come a point, and I think that there will come a point where society is actually going to need to have its butt kicked. Is There' be a point where actually driver? Those cars do make sense, and there'll be some weird accidents, and people will be like. I don't want these drivers   209 01:01:05.010 --> 01:01:20.679 Professor Marcus: things, and then we will at some point be a net benefit. It will actually be worth it for us to have those cars, even though there are those weird accidents, because they will cut down so many others. I just don't think that we're there yet not sure would be there in the next ten years. But i'm sure we will be in the next century.   210 01:01:21.140 --> 01:01:41.000 Professor Provost: So uh, unfortunately our time up, that went very that went very fast. Uh, thank you very much, Gary, for you know, for for for joining us. Thank you, everyone for for for for being here. Uh stay tuned for other Fubon uh Fubon center uh the center events. Uh and yeah again, I really appreciate it. Uh, thanks, Gary,