**A149C280\_2310050L\_C3005.mp3**

[00:00:00] **Speaker 1** Awesome. Yeah. Okay. Should we wait for up to you?

[00:00:06] **Speaker 2** We could. Yeah.

[00:00:07] **Speaker 1** Oh, she just texted. Awesome. Um, I guess we can. We can get this started, if that's okay with you.

[00:00:32] **Speaker 2** Sure. Um. Cool.

[00:00:38] **Speaker 1** Already? Yes. Yeah. Well, maybe so. Yeah. Okay. Awesome. Perfect timing. Thank you so much for sitting down talking to me.

[00:00:56] **Speaker 2** Yeah, Thank you for having me. Pleasure.

[00:00:58] **Speaker 1** Um, I guess my first kind of question is, how would you define it?

[00:01:03] **Speaker 2** So, Jane, I would define a I kind of broadly as the collection of systems that help with predictive capabilities and making decisions. So historically, there have been two ways of approaching this. One is through machine learning, which has mainly been using pattern recognition to predict what may happen in the future. The second was sort of more what's called rule based of trying to mimic kind of deductive reasoning that people use. Nowadays, most things fall into the machine learning category, especially some of the technology that we see out there today like CBT and those kinds of things.

[00:01:47] **Speaker 1** Is that was primarily used in classrooms and universities that students use.

[00:01:52] **Speaker 2** So. Uh, depends on the context. Right. So in research, a lot of the focus has been on developing new techniques for machine learning and how we can extract insights from data and forming prediction. But the systems that are used from students are the products of these kinds of methods. So, for instance, using CHIP to help with writing, um, use these methods to create the kind of language model that's able to do word prediction. Same thing with, you know, using next word prediction on your phone. That's the mechanism behind it kind of makes it operate.

[00:02:29] **Speaker 1** And what kind of information or where do programs like GPT pull their information or data from?

[00:02:37] **Speaker 2** So, um, it's a great question. It really depends on sort of the goal of the particular program. So with chat, GPT specifically used the corpus, right? So all kind of written word from the internet from I believe around 2019 or 2021. So it doesn't have any more up to date information than around that here. But that's still a ton of information and it's every single kind of written word and every single written article that open that I was able to access after that point that it used to train a model to figure out its language. But for example, other models that might do, you know, image recognition or something like mid journey that takes prompts and then turns them into photos, they need to have access to photo databases. So that's going to be completely different from the way you'd access or the written word database and depending on the application might be something. So some of the stuff I work on, for instance, in like medical applications, I go to like the hospitals or I go to medical collaborators and go through those approval processes to make sure that I have like the relevant chart data or relevant MRI's and all that kind of thing in order to form predictive models.

[00:03:50] **Speaker 1** So a lot of that information is still relatively new if it's only from 2019. So what happens when the information that is coming from is misinformation or information that is incorrect or accurate?

[00:04:01] **Speaker 2** So that's a great question and that's a big concern. Um, but I think it's important to think about what is the goal of a system like CBT. So the goal of the system isn't necessarily to provide information in the form. It's not a search engine, right? Like we go and Google, we put in our query, we're looking for an answer to a specific question and we're hoping that we get recommended, you know, relevant articles. Judge CBT is designed to give responses that fit well with the queries we provide it. So it might not necessarily so Whether the information is correct or not might not necessarily be as big of an issue in the way it formulates its answer, as would someone give an answer to a question that looks like this. Now, the way that opening, I obviously deployed it. There's a lot of other different engineering kind of wizardry that goes around it to make sure that it limits the amount of misinformation that it provides. Right. That the caveats like a lot of answers, saying like, oh, this is might be beyond my knowledge. Um, but really the way to think about it is the system tries to produce an output that sounds good. So we said it limits itself for 2019. You could ask it, you know, the results of like let's say the 2022 midterm elections, even though it doesn't have access to that information, it'll give you some very plausibly sounding good response about how those elections went, even though for sure it has no idea you can even asking for what's the weather outside. It'll give you a response where, oh, this is if someone were responding to the question, what's the weather like outside? This would be a response that someone could give, but it might have zero bearing on what's actually happening outside.

[00:05:41] **Speaker 1** Dr.. So would you say it depends really on the user and what they want to do with the information and how they want to use the programs.

[00:05:47] **Speaker 2** So I would say a lot of it has to depend on the user. I think that's a very good observation, but I think the it's incumbent on the user to realize some of the limitations of what the program's meant to do and what the task you could do with it is. But there is I think this is actually like in the past week or the past two weeks, there's been some recent developments where they've enabled real time data collection to interface with with systems like CBT. So for instance, like being able to give it like camera sensors and then describe the room and like what's inside the room. So in principle, right, we could design these systems in the very near future with immediate sort of web access and maybe allow them to like set queries to Google to find more relative information, then compile that into answers. We're not quite there yet. And and so right now. Right. It's understanding that the information is limited and what the tests that we can actually perform are with that.

[00:06:47] **Speaker 1** So programs like Touch and Beauty are relatively new, only being released last year. Was there any reason for being. At a particular time like last year or what was kind of the timeline of these programs are coming out.

[00:06:59] **Speaker 2** So. Right. Like science advances, is it? That's right. We need the right kind of technology. We need the right compute. So a lot of things had to come together to have something like to actually work out. So large language models have been around for a while in terms which is like, uh, probably like, you know, seven, six years. But this idea of taking large amounts of language, turning them into representations that computers can understand and then using them for prediction tasks, we've had some component of that for a while. So there have been lots of systems, for instance, trying to figure out if I look at movie reviews and I want to figure out, do most critics like or dislike the movie? So this is kind of a test called sentiment analysis that uses similar kind of language processing. But notice the six data and processes it. It doesn't generate new responses. There's a whole other side of things, which is the generative model part which needed to advance of how I take models and then use them to reproduce kind of new data that seems reasonable. So if you remember, even this kind of stuff has been around for a while. We have these deep dream kind of pictures, right, where you took, you know, a classic painting. And he said, Right, repaint this with only dog faces and you have the thing inserting the dog face. This was like the beginning of generative models. So what sort of had to happen was a convergence of these two fields of the large language models, the generative models coming together, and a lot of clever engineering work to make it so that this can actually be deployed with the given computer resources to make it accessible. But I think the biggest thing that Chelsea had going for it, because GPT models have been around for a while, so by the GPT that's currently free is GPT 3.5 and Openai has GPT four, which means it's the fourth version of the gypsy model. But uh, so let me get back on track here. Even though these these things have been here for a while. Losing my my spot here for a second. With what? With what? I was going with this.

[00:09:22] **Speaker 1** Uh, yeah, no worries.

[00:09:24] **Speaker 2** Yeah. Yeah. Can you rephrase the question? Yeah. Sorry.

[00:09:26] **Speaker 1** Um, was there kind of a reason that this program.

[00:09:29] **Speaker 2** Oh, yeah, I think so. So. Right. So the the big thing, the change was actually the chat part. So the fact that there's a user interface where you can interact with natural language with a GPT model that sort of made it really accessible, that kind of democratize access to to GPT. So even though GPT 3123404 is only been around recently by 123 have been around for a while. Really you can only deal with them if you had like you know, a computer science degree and like wanted to give command line input. But the fact that there is a user interface that allows you to chat is what has I think invigorated interest in, in that design work really kind of elevated it.

[00:10:07] **Speaker 1** Right. Absolutely. So what is that experience been like for you with your students? How to students use either in the classroom or out of the classroom for assignments in your classes? I students use em.

[00:10:19] **Speaker 2** So, uh, sometimes they use the AI because I assign them to use AI. So like when I teach by supply chain class, for example, and I asked them to do like forecasting for like the demand, they have to use these machine learning and A.I. tools in order to do forecasting. But something has been really useful. Is using GPT three for editing different types of documents, be it you know, I have students, for example, that speak English as a second language, like I myself speak English as a second language. There's some things that don't come naturally if you're not a native speaker. GPT comes super handy if you give it, you know, a prompt like, Hey, can you edit this to make it sound like more natural or can you edit this to make it sound kind of more concise, Feel like this is over long and it's been really, really great in basically shortening the amount of times of like writing through papers or even writing emails in responses, making them sound more professional and assisting with communication that way. Those are the two kind of big things I see.

[00:11:18] **Speaker 1** Absolutely. Um. Kind of going with that. How, in your experience, have students used A.I. for academic misconduct or using A.I. to kind of write essays for them when that wasn't particularly be assigned?

[00:11:32] **Speaker 2** So I haven't seen that yet. I don't I don't assign a lot of essays in my classes. It's engineering class and thankfully, well, not thankfully. I guess we'll see whenever this bridge is a gift. But like the. Tragedy style models. See, some models in general are not very good at mathematical reasoning, which is very interesting. So they can reproduce grammar very well. But if you start asking it to solve different math problems or provide you with proofs for different mathematical propositions, you're going to get probably nonsense out of it unless it's something like very, very simple that's been reproduced thousands and thousands of times. So we're not quite there yet. But something that would be interesting in the future is something we've discussed here in the department is, is we teach, you know, both undergrads and Ph.D. students. We're supposed to be like future experts in these things instead of just asking them to reproduce sort of classic results or learn clinical classic material, presenting them with a proof or a paper that's been created using a generative model and asking them to find kind of errors in that paper, because that's probably going to be where a lot of impact comes in the future is as this stuff gets more common, gets sent to peer reviewed journal. So we need reviewers and we need scientists that are on the watch to make sure that things are up to snuff.

[00:12:57] **Speaker 1** So in your opinion, what's the best way to move forward to integrate A.I. safely in whatever it's being used for, whether that's the classroom or for personal?

[00:13:07] **Speaker 2** So I do. I think the best way to think about it is there's not one single standard for safety. I think the important thing to consider is depending on the situation, depending on context, who the system's interacting with and who's deploying it, what is the measure of safety in the classroom? I think we have a relatively straight, more or less straightforward bar to clear here. If we use it in terms of language editing tools or use it to aid with like word processing, I think that's great. Right. The introduction of Microsoft Word didn't diminish academic work from, you know, writing things by hand. And I think this is just kind of the next step of that. But other concerns that some folks might have may be in terms of, you know, less savory language or including different kind of biases and assertions. That's something that we got to be more on the watch for. And that's just something that you have to sort of inform the students about in occurring. And if anything, the incorporation of these tools might actually aid in detecting and stopping plagiarism, because there's a strong tendency of chatbots to insert citations that don't exist into work. So it's very easy to find misconduct for students towards signing papers that that don't exist, but that might also make someone a bit more cognizant of like, oh, okay, this is why I need to make sure that I actually read through this and operate thoughtfully.

[00:14:44] **Speaker 1** Right. So how can people who help design or influence programs such as me to make sure that the information that the program is pulling from is correct and accurate? Or what can they do to make sure that steps of misinformation don't happen?

[00:15:00] **Speaker 2** I think there's two different levels of the misinformation, right? The one level is making sure the information comes in is accurate, and that's difficult. So if you're trying to learn language, the more examples of language you have, the better. And so you're not necessarily going to just moderate the content, right? Just because something is fake news doesn't mean it's not written in proper English. So that's more challenging. But the other side that's more actionable and needs to be watch out for is that because GPT three produces good sounding answers, it does so at the level of like a college educated individual. So we have this concept called like the Turing Test, which is if you interact with like an A.I. model, can you discern the difference between it being artificial intelligence versus it being human intelligence? And so previous systems that have kind of passed this have been at the bar of like communicating as if you were like a 13 year old or like a teenager and just communicates at the level of someone with an undergrad degree. So unless you're probably a somewhat of an expert in what you're asking it about, a lot of it sounds very reasonable. Even though it might be completely off. And that's the really challenging aspect. So I don't think so. One thing that openness has done, which I think is good, is make. Kind of hard coded decisions such that GPT or the churchy party version caveats a lot of what it's saying and I think that's a good initial step, but I think providing like an annotation system or providing some kind of other process or more warnings or also public education about the limitations of these systems might be what we need to avoid just someone relying completely on generative information versus even going to like a search engine afterwards to fact check what it is they're seeing.

[00:17:02] **Speaker 1** Right. And would you say the people who are more hesitant to using AI. Generally in either the daily life or in schools or people who are just kind of resistant to using it.

[00:17:15] **Speaker 2** Resistant. So. I'm not sure. I think it's it's going to be difficult to avoid. I think there's too many good use cases of AI and different machine learning tools. It's going to be hard. It's. It's going too fast. Approach the level of avoiding using a computer, avoiding using the Internet, avoiding using electricity, everything from the way we interact with like social media on our phones and everything from like the mail being sent to us. The way our credit scores are determined. If you want to interact with any sort of institution, unfortunately, you're going to have to interact with. Right. So I think it's more of a making peace with the fact that it's there and trying to figure out maybe ways of being informed, being involved, and kind of lobbying for ways of having additional ways of expressing both like dissent and expressing concerns to aid in like the design of future models and making sure that you have a stake in in the way things evolve going forward.

[00:18:31] **Speaker 1** And so what do you think the future of AI looks like right now in the in the near future, in a six month period versus five or ten years?

[00:18:39] **Speaker 2** Okay. So this is like always was like a dangerous to make to to a far reaching conclusion. I mean, the next six months will see the advances have been really, really strong with the way these GPT models have been going, especially because it seems that just by increasing the size of the model and increasing the amount of data that we use for training, that improves the capability of reproducing language very, very well. But I think in the near term, what we're going to see is a lot of specializations of generative models. We're going to try and see them being deployed in various aspects of think of like systems that might annotate doctors notes in the medical field. So some kind of generative model that specialized for like medications or generative models that are specialized for, you know, white collar office workers to write quarterly reports, models that might be used. You know, I don't want to you know, journalism is important, right? But think of like standard kind of articles that might be publishable sports results or like summarizing things from like Capitol Hill. That might be something that is being worked out. I know that those systems have already been in development for a while, but something like that is going to probably take more hold. Same goes on in ten years. Like, who knows? I think the field has been moving really, really fast. Some things that have seemed promising ended up being put on the wayside. Some things that didn't seem promising at all came back like even neural networks, which are sort of like the basic model that underlies a lot of modern day machine learning. They weren't invented in the eighties, and then they were kind of put on the back burner for almost 30 years before we finally had enough computation problems, cavitation power to bring it back and reform the field. So it's very hard to say what's going to win out in the long term, but I think we'll see a lot of these specialized tools. And I think the other thing we'll see is a lot more discussion on what regulation of AI looks like and what is the WHO ownership of data and the relationship between the private consumer, as well as the companies and institutions that use these kinds of tools.

[00:20:51] **Speaker 1** Is there anything else you'd like to add about A.I. or any misconceptions that you'd want to clear up that general generally people have about? Yeah.

[00:20:58] **Speaker 2** Yeah. I think the most important thing is like we use terms like artificial intelligence and it sounds kind of scary, right? Like the image that I conjure up is like, you know, how 9000 or like Terminators, you know, crazy stuff like that? I think the important thing to remember at the end of the day, it's not it's not truly intelligent at the moment, as far as we know. These are just very capable tools that are useful, that are designed with a purpose. They're designed by people and they have limitations. It's just kind of like the next step in helping us improve, right? Kind of like how we replace horsepower with machine power. We replace, you know, writing things on clay tablets with word processors and, well, typewriters, then word processors. This is simply the next step of taking some of these additional tests that, you know, required us to to do before more manually that might have been more menial. We can now automated it's new tools that enable us to do that.

[00:22:06] **Speaker 1** Awesome. Thank you. I think if I have any more questions, is there anything else you want to ask more questions in? I don't think so. The founder of. Open at. Testifying before Congress about. The key to our essential.

[00:22:25] **Speaker 2** Yeah.

[00:22:27] **Speaker 1** Is that hyperbole?

[00:22:29] **Speaker 2** So I would, I would ere on the side of. Yes, but but I think it's let's let's a caveat caveated Right. Could it lead to our eventual demise. There is there is a possibility where that could happen. But I think looking at what might happen in like 100 or a thousand years is important. I think there's like a few more close term ethical concerns that we need to think about. So, for example, if we start having integration of autonomous vehicles, right, what are the justifications we're going to have for preserving the life of the driver versus preserving lives of pedestrians or other people's walk by as we deploy AI and prediction in the medical field? What are the ramifications of having diagnoses made by automated systems? How does that look like when you integrate it to doctors? Whose responsibility is it? If the prediction is correct and then the doctor decides to not go with the AI prediction? Right. These these things are a lot more immediate. The other thing that's also a bit concerning to me is in the same testimony, the founder of Openai, I was talking about how there's a need for more regulation in the slowing down of the process. And I think it's kind of important to note that he started kind of this this feels like a little bit like potential regulatory capture in the sense that as soon as GPT proves to be the generative model that kind of wins out in the market, that's when open AI is more clamoring for that. But it's obviously hard to say, right? Like what's his is going in there? Like there are a lot of like legitimate concerns, but still weighing the democratization of like developing this technology versus the safety risks of making it too open ended. That's something that's going to have to be deliberated and that's something that's going to have to discuss that's a lot more near-term than the existential risk aspect.

[00:24:31] **Speaker 1** Is there such a thing as going to.

[00:24:33] **Speaker 2** Is there such a thing as going too fast? Uh. I guess. I guess there might be, but. It's hard to judge the speed of advancement, like what's incremental versus like, what's a paradigm shift? Right. You don't know that you've gone too fast until you've actually gotten their rates. It's hard to to see in between. Like I said, right, to get to Chad, having the global influence and in causing, you know, everyone to be interested in. I there had to be hundreds of small advancements along the way before we got there. Right. No one cared, really, other than some people on Reddit and some people online about GP2 and like the employees that opened the eye. But now the GP 3.5 has a chat feature. All of a sudden we realize where things are so hard it's hard to judge.

[00:25:32] **Speaker 1** I think it's everything. Thank you so much for speaking.

[00:25:36] **Speaker 2** Yeah, of course. No problem.