MIT CSAIL Alliances | Singh Bobu Final Show

Welcome to MIT's *ComputerScienceandArtificialIntelligenceLabsAlliances* podcast. I'm Kara Miller.

[MUSIC PLAYING]

On today's show, AI has the potential to transform health care, but it may be a bumpy ride.

On the research side of things, we have really carefully researched tools that never get implemented. On the vendor side of or industry side of things, there are tons of things that are implemented, and we have no idea whether they work.

Dr. Karandeep Singh, Chief Health AI Officer at the University of California, San Diego, is here to explain what's working, what's not, and what the reality is on the ground.

I think people look at my title and role as a Chief Health AI Officer and think that I'm going to be the cheerleader for all things AI. And I would say, the more you know about something, the less exciting it becomes and the more you appreciate what the challenges are going to be and what things just simply aren't going to work, not necessarily for technology reasons, but for social reasons.

Plus, a peek into the lap of MIT researcher Andreea Bobu

Language is a very natural form of communication for us humans. So what this means is that we can communicate back to our robots what we need or what we'd like, what we don't like in a much more natural way.

That's all coming up in just a minute. But first, MIT, CSAIL and MIT xPRO are offering two new online professional programs this spring. One is focused on what technical leaders should know about cybersecurity. The other is designed to help companies apply AI to their business.

Find more info on our website cap.csail.mit.edu You can also email us, podcast@csailmit.edu. Listeners to the podcast get 10% off the course. So, again, the website, cap.csail.mit.edu and the email, podcast@csailmit.edu.

[MUSIC PLAYING]

Going back decades, medicine has tried to use data and information to get clues about the future. When you go into the doctor's office, you get asked a lot of questions, in part, to try to understand what might lie ahead. And AI could mark the next data revolution in medicine. But--

What we have is this interesting dichotomy that's emerged in health care tech, which is that at the same time as we have AI being involved in supporting medical decisions, we also have fax machines and reliance on old infrastructure to provide a backstop for basic health IT and operations.

So the transition to more and more AI, it's going to be a big one for health care. "But to some degree," says Karandeep Singh, "AI has already crept into the system in at least a couple of ways."

So a number of hospitals around the country have implemented what are called deterioration models, which are essentially AI tools that we use in the hospital to detect when a person is getting sicker so that we can go and check in on them and make sure we correct their course. And the other, I would say commonplace, which is related is sepsis, which is a body's reaction to severe infection. And similarly, when sepsis comes on, it can progress very quickly. And so we have tools that we use in the hospital that are AI tools to detect sepsis and to direct our clinical attention to it so we can get it managed appropriately in a time-sensitive way.

Now, there are opportunities for AI to potentially improve care, but testing AI in a hospital or a doctor's office isn't easy. Plus, there are other complications. Years ago, I talked with a doctor who helped run a hospital. And he said, "Doctors and nurses were so overwhelmed by electronic alerts, it was hard for them to actually know who to help first." They were getting buzzed and beeped all the time.

So technical interventions are going to have to avoid that sort of alert fatigue. You can't overwhelm the people who actually deliver the care. Singh, who is also a Professor of Digital Health Innovation at UC San Diego, says, it's crucial for those who create AI models to answer questions, like, do people actually get better care when you're working with these models? Do fewer people die as a result of implementing them? And then there's a final critical question, can you convince doctors that they work?

For pills that we take, there's a pretty robust requirement that you have to show some kind of clinical endpoint getting better for that drug to get approved and end up on the market. If you look at the FDA process right now for clearing software as a medical device, when it comes to AI, there's no requirement that you have to show clinical benefit. You really can ride mostly on demonstrating accuracy.

And so what that means is one of the challenges actually is that for any given outcome I might want to predict, there are dozens of companies and startups in that space doing that prediction. There are dozens of published papers showing that you can accurately predict the outcome in some cases. And yet, we have just not enough evidence to show what's the right intervention to pair to it. And to your point, is alerting the actual appropriate intervention?

A lot of the recent papers I've been seeing that have been showing benefits of I have not used alerting, but rather have used sorting of worklists, where you don't alert clinicians. What you do is you sort their work in ways where they would have gone to see another patient next. But now they direct their attention to the patient kind of most in need on their existing work queue next as a result of the AI model.

So, is there a very robust industry of startups or big companies out there that do try to sell hospital systems or doctor's offices on you should use this?

Yes, there are lots of people. And it's not actually only external companies, it is the electronic health record vendors. Electronic health records are essentially our go-to tools for delivering and documenting and billing for the care that is delivered within health care environments.

And so that's like the central software engine, the kind of operating system that powers American health care. And the electronic health record vendors themselves have dozens and dozens of AI tools that you can turn on at any point. Separate from that, there are small and large startups and big companies in this space developing and disseminating models.

So, I think, there's no shortage of models, if you're on the health system side of things. What there is potentially a shortage of is knowing when should you use which one, because of the fact that there's very little outcomes data showing that something works. And I kind of refer to this idea as the health AI paradox.

On the research side of things, we have really carefully researched tools that never get implemented. On the vendor side of or industry side of things, there are tons of things that are implemented, and we have no idea whether they work. And so I think both things are happening at once. And it's a real challenge for a health system to decide which course to go when the tools that are available to implement are completely separate lists from the set of tools that have high quality of evidence in the scientific literature.

So how do we get to the place or are we getting to the place of what you talked about with drugs, where, as you say, when a drug is tested ideally, you've got like a randomized controlled trial. You've got some people taking it, some people not, people don't know. You're trying to say, everything being equal, how's this drug working against not taking it? Are we anywhere close to that with AI?

Yeah, I think, the reason it's been tough to do with AI is that we know that models don't always transport between settings. So a model that works well at one hospital may not work well at another. And similarly on outpatient or people at home side, a model that works well in one population may not always work well in another. And as a result of that, it's really hard to set up the same kind of trials that we set up with drugs when the actual tool itself requires a lot of configuration and coordination across a number of teams. It often also requires prioritization from a health care leadership, operations standpoint.

So the challenge is that whereas with drug trials, a researcher can really set up a clinical trial and, with a drug, do a placebo controlled trial with coordination with a pharmaceutical company, to do an AI trial where you actually change an operational clinical workflow, you can imagine that that's orders of magnitude more complex because you actually can't have a researcher by themselves drive that. It needs tight coordination with health system leadership and operations. And that's where I think you've seen the larger systems that have consolidated leadership and processes across multiple hospitals be the ones that have taken some of the lead on doing these outcome trials because they have tighter coordination between operations and research.

You mentioned something that, I think, is really interesting, which is that data in one place is not necessarily portable. And so I wonder if that's a problem where either a company or a hospital chain will say, well, we figured out this great AI model. We used all our hospitals in Connecticut.

You should totally use it in New Mexico. Only problem is different demographic makeup, different income levels. And I don't know, maybe you cannot graft a bunch of Connecticut models that work onto New Mexico.

You've just described the modern state of health care AI, which is precisely that. Companies come to you in some cases with a lot of high-quality data from a system that potentially resembles yours and potentially doesn't. And the challenge is it's not just a demographic thing. This is not a matter of the model works in this patient population, but not in this other one. It's also data quality.

And the way that your clinical care gets documented and generates that digital footprint, that process might be slightly different at one hospital or one health system versus another, and that could lead to big differences. Which electronic health record you have might greatly shape what that data looks like. Which laboratory you use and what some of clinical care things are might be different, what the processes are.

At one hospital, a referral to a specific service or being on an ICU floor might have a specific meaning. It might mean that you need frequent monitoring of your sugar. At another hospital, they can do that sort of monitoring on the floor.

So even important contextual clues, like where is the patient located, are they in the ICU or on the floor can mean a very different thing in two different health systems with different acuity levels. So I think that the issue is that you don't really know until you look. And there's many ways that it could work out and that the model just works. But there's many, many more ways that it could potentially fail, which is why there is this early discovery period where you have to try to run these tools in the background or with historical data to see, is the information that they've collected from other health systems actually relevant to yours?

You have talked before about the medical record software provider Epic. And they created this-- I found this to be a fascinating--story. They created a no-show predictor, basically, who might make an appointment but then not show up for their appointment.

And this could be useful, you could imagine, for a doctor's office or whatever. But as you pointed out, there were some issues with their model. Do you want to talk about that?

Yeah. So we've looked at a couple of different Epic models. I think with the Epic No-Show model, we haven't published on that. But my colleagues at UCSF, including Bob Wachter and Sara Murray have.

And some of the issues with that model were that in the original version of that model, there were some sensitive attributes included about people's demographics that were driving the prediction. And we don't know how important those variables were. I don't want to misquote it, but I think ethnicity was one of them and potentially religion was another.

So those are things that you have to ask yourself, why would we want a prediction that accounts for those things? It's not entirely clear. Now, to Epic's credit, they actually got rid of some of those things when they went to the version 2 of their model.

And they also clarified in their documentation that whereas originally, it wasn't clear what the appropriate use of that model was, they clarified that this model should not be used to double book patients because that could produce inequitable reactions to no shows, which is that two people show up who require a lot of time and attention, but actually were folks who were predicted not to show up because of their social determinants of health, essentially factors like poor transportation and complex medical issues.

So I think there's a-- the cautionary tale there really is that you have to be really transparent about what information these tools use. Just because a lot of other health systems have adopted them doesn't necessarily mean that, like everyone's done their due diligence. And I think, Epic and many other vendors have learned from their mistakes on some of these things.

And so now, like I said, their guidance is pretty clear. But that was a place where, I think, there were a lot of alarms raised around, is this the right way to use this tool? You can't just give us a tool. You have to give us a appropriate way to use this tool in a clinical workflow that actually makes sense and is equitable.

So there's a lot of folks who, I think, are very excited about the potential of AI to transform health care. When you hear that level of excitement-- you're very much on the ground and you see all the shades of gray here. I wonder how you think about this.

So I think one of the funny things to me is that, I think, people look at my title and role as a Chief Health AI Officer and think that I'm going to be the cheerleader for all things AI. And I would say, the more you know about something, the less exciting it becomes and the more you appreciate what the challenges are going to be and what things just simply aren't going to work, not necessarily for technology reasons, but for social reasons and other kind of norms that we have that we would have to change over time. So I actually think that a lot of times when there's excitement, you have to channel that excitement and try to figure out what is the excitement tangibly about.

In some cases, I think, the excitement is really, really, really warranted. I think the excitement, for example, around ambient documentation. The ability for a clinician to walk into a room with a patient, ask their permission, record that visit with audio, and have it generate a draft of a clinical note 30 to 40 seconds after the patient walks out of that room is something that is transformative for doctors and nurses quality of life potentially.

Right. You don't have to sit there and type it up or go home at 10:00 at night and do it and whatever.

You still have to edit it, and you still have to look at it to make sure it's accurate. But right now, there's a whole industry around providing that kind of support where no one can do it in 30 to 40 seconds where you can do it between patients as an example. So I think that's one where you understand that's a real pain point, and the technology actually directly addresses the key part of the pain point.

On the other hand, though, there's a lot of times where people will say, let's leverage AI for X, Y, Z, and I think it's almost the opposite. Define X, Y, Z, and understand what is AI going to actually do for you. Predictive AI can give you a risk estimate. What you do with it is essentially the benefit that you're going to see. And unless you have a really good plan for what to do with it, the best prediction in the world isn't going to help you.

Let's talk about that a little bit, because I do feel like at least in the last several years, there has been some excitement about Watson from IBM and these ideas of, well, an individual person is not going to have read every paper about this type of cancer or know that Japanese women are more likely or this or that. There's all these subpopulations and all these different things going on. And so, how helpful are these models at this point at helping you or people to predict this is the ailment you have or this is the likelihood of this type of cancer or that this type of cancer is going to lead to X, Y, Z? I mean, I just wonder where we are with those.

I think where we've really progressed in the past, one or two years is going from a world where answering that kind of question required a really bespoke prediction model that you built very carefully by hand that could do that one thing. And the challenge was is that you want to do something slightly different, you had to go back to the drawing board and train a new model to do that slightly different thing.

With the rise of foundation models, which are the underpinnings of large language models, ChatGPT, Llama 3.2, and some of these open large language models, what we're seeing now is that you don't necessarily have to go back to the drawing board each time. You can borrow some global knowledge, which is imperfect and biased in some ways, but captures really a large understanding of a diverse set of concepts. And you can now take that knowledge and as a starting point, which means that you can get to reasonable predictions with a lot less training data because there's already some information about the world encoded in these kinds of foundation models. So I think when we talk about prediction, where I really see the big advancement is that we might be able to do better predictions with less sample size because we're starting with foundation models as a starting point rather than starting from scratch.

But the sort of things that you mentioned around excitement around Watson, to me, that actually is a more of a search and question answering capability that really has matured a lot in the past couple of years as a result of large language models. And I think there when people have a lot of excitement, the excitement is because they previously had to be a domain expert and understand a bunch of programming to get something done or follow a recipe where they did a bunch of different steps. And now they can just explain the recipe in a prompt, use a large language model that's, in some cases, connected to the internet or other high quality sources, and without knowing really much more beyond just English, they can get to information that actually answers questions in a way that they never could before.

Now, that information might have problems in it. And it's not perfect, and a lot of research happening in how to make that information better. But I think that actually is a key central place where the field is miles ahead of where it was two years ago.

And I think my advice to clinicians to manage the excitement there is learn prompt frameworks. Most of our end users don't know what a prompt framework is. And what I've been trying to--

Are the end users-- the end users are the doctors.

In a lot of cases, they're the doctors. I would say our patients actually probably know how to use large language models better than our doctors do because patients are usually the first to try new things well before our clinician workforce does. So I think the gap isn't so much our patients don't know how to use stuff. It's actually probably more that our doctors don't know how to use it and our nurses and our clinical workforce.

And the ones that do, I think, aren't being given a way to share that knowledge. So I think, to me, democratizing and training up our front-line workforce on how to effectively use large language models and understand what are the constraints of one tool versus the other tool based on some really basic key functionality differences between language models is something that, I think, we really need to be investing in. And that's where a lot of my kind of roadshow internally has been, is in getting people up to speed on the current state of large language models and how to use them to do really basic information extraction tasks of this sort that take a lot of time right now to do, but that we could get additional support with using these tools.

So, for me as a patient, give me a sense of how doctors access to some of those tools that we talked about in the last few years, predictive or ability to ask questions, how does that change my life? How does that improve things for me as a patient?

So I'll give you a couple of examples of this where it's touching patients directly. Let's look at our sepsis model that we have at UC San Diego Health, which is called Composer. So this is a model that's deployed in our emergency departments, and it's a model that monitors people on a really frequent basis once they arrive in our emergency department and generates ongoing predictions for their risk of sepsis.

And when their risk exceeds a certain threshold, then the model applies a whole bunch of logic to decide whether an alert should be generated or not, and then generates an alert to the clinician when it feels like that clinician hasn't recognized sepsis or there's something missing that the clinician needs to be taking a look at. When we've actually implemented that, that resulted in a nearly 20% relative reduction in mortality. So fewer people are dying of sepsis in the hospital than were before.

And it's to the tune of something like 50 patients per year fewer dying across the health system from sepsis than were before we deploy that tool as part of this workflow. So what I would say is that's a place where, again, it's touching the lives of patients where there may still be alert fatigue. It's not a perfect tool, but where we actually have some measured benefit through a deployment where we were able to track changes over time.

Another example I would say is in some of the work related to in-basket reply. This is where you message a clinician's office. that message might go to a nurse, a triage nurse, or it might go to a clinician. And a clinician will need to take some time to draft that reply and think about what they're going to say.

And some of that is medical advice. And some of that is, like, hey, how are you doing? It's great to hear from you. And so one of the things that we were able to do here is look at, well, does this actually save clinicians time?

This tool actually is not meant to save patients time. It's really meant to save clinicians time while still maintaining appropriate response back to patients. But when we deploy this and we studied it in a randomized trial, we found that this tool actually didn't save clinicians time. It actually increased the amount of time they spent when they used it because they spent more time editing the message than it would have taken them to draft it in the first place.

However, clinicians seem to like it. And if you look at other studies that have been published out of Stanford, it looks like that it actually reduces their cognitive burden. So, I think, we're starting to understand, again, which of these things save us time? Which of these things improve quality of life?

And even when they do save time, do they save time during work hours? Or do they save times after work hours? Because the implications of that will determine whether this is a tool that increases productivity or whether this is a tool that improves quality of life.

So those are two examples where I would say, use of the tools touch patients. And that's why I think for the in-basket reply, we've been on the cutting edge of making sure that we label all of our auto-drafted replies after, even if they're edited by clinicians, with a postscript that tells people, this text was partially automatically generated and then reviewed and edited by this following clinician. So that's where, I think, yes, we're trying out new things, but I think we're doing it in a transparent way and studying the impact of that and discovering when it's not impactful, that we need to change course.

So look down the road for 5 or 10 years. How could you imagine, in an optimistic more optimal scenario, how would AI change life in the hospital, if you felt like, yeah, I mean, this is like, this is the hopeful scenario?

So I think we have a lot of AI that's really, really patient centric in the sense that the patient that's in front of you then and there is the one that we're using the AI tools to connect with and reach out to. One thing, I think, we really haven't done is use AI at the system level to coordinate care viewing our health system as a true health system and not just a series of health care encounters. So I think in the next 10 years, a lot of the emphasis is going to shift away from the hospital, even away from ambulatory care and into the home setting where we can direct people to care when they actually need it rather than having them wait.

I think to make sure that when they do need care, they get the right care, I think, we will have better tools to forecast when we're going to have shortages or longer wait times at one hospital versus another. So we can direct people to the right level of care and potentially to the right facility when we have multi-hospital, multi-clinic health systems where we can load balance a little bit better than, I think, we currently do. You can imagine within the same health system, one setting might be busy and the other one might not be. And yet right now, we have no way to really redirect people in a way that gets them better wait times. So, I think, from an experience standpoint, my hope is that the technology will help people more at home than necessarily in the hospital and that the technology will help people get care faster and with a better experience by virtue of understanding the system more at a holistic level than just purely at the encounter level.

When you talk about helping people at home, I don't know if you mean partially in the predictive context of like, I don't even know if this would factor into an Apple watch, but basically trying to pick up signs of problems before they developed into a problem that was so big it required surgery, major intervention, something very expensive, that sort of thing.

I mean, in both the predictive and generative context. So, I think, in the predictive context, we can predict all sorts of events on the outpatient side. What we don't have is a workforce that is proactively calling people at home in response to signs generated out of predictive tools. That workforce doesn't currently exist.

I'm not saying that the health care system will expand to add that workforce, but I do think that there will be shifting of priorities where that starts to become more of a priority as some of the other things start to mature and get addressed. So I do think we're going to see shifting of workforce roles into that in the next 10 years or so. And I also think in the generative context.

I mean, right now, people can hop on public ChatGPT, public Claude and ask all kinds of questions about their health. And right now, there's a big gap between what people can get there versus being able to schedule an appointment through a text message interaction or get some basic health questions triaged and answered all through a chat-based interface. So, I think, we're going to be seeing more community and patient-facing chat bots where people will be able to get things triaged a lot sooner, hopefully, and in a lot of more of a low touch way where you're not waiting on hold, but you actually can get some initial concerns dealt with without having to get high touch care.

Almost like in the way that Blockbuster introduced friction to get movies that was removed with the introduction of Netflix, I think, we'll start to see more of the friction of getting clinical care start to gradually get offloaded as we figure out what are those things that we already follow a decision tree that we can really start to delegate more of that to our patients through a process that relies in part on collecting information through chat bots.

And when you think about people maybe knowing that something more serious is coming before the heart attack happens, who do you imagine would be the person or the entity that would be like, sir, I think, like, these signs point to X happening? You might want to get to a hospital before something worse-- before something really bad does happen. Is that a new industry that's going to rise up? Who would be doing that?

It may be a new industry, but I think the way we have things set up today is that people have a medical home, and that medical home is often linked to your primary care doctor. And so, I think, right now, if you were to ask me that question, whose responsibility is it, it's some combination of their primary care physician and the population health office within a health system. And so I think that we'll have to figure out how do you do this in a way that doesn't double, triple the workload for our primary care workforce.

They have enough of a hard time managing just simply all the messages they get and the people that are in front of them and seeking to see them and appointments day-to-day. How do you expand that reach even further without burdening those folks? And that's why I think that regardless of where responsibility sits, there's going to need to be an additional almost like workforce whose job is to manage some of the proactive things like that.

In the case of that, though, I mean in the case of someone risk of heart attack, it may be their primary care physician. It may be their cardiologist. So I don't think we need to invent new rules where we have existing folks with capabilities and responsibilities. But I think we do need to give those folks digital support that makes it possible for them to have a broader reach than they currently have.

So you drew out a pretty interesting and positive scenario for what AI could add to the health system, things that are really needed. What are your biggest concerns in the next 5 or 10 years about how AI will unfold and issues that may crop up?

I think there have been a lot of cautionary tales in the past several years for AI and potential downstream consequences to patients that weren't anticipated or weren't planned for. So, I think, my biggest fear is that we will have a lot of AI that is getting deployed, but no one will know does it actually work. And I think in the short run, the excitement around AI might be enough to keep that kind of an industry afloat.

In the long run, though, when budgets get tight and people want to know what actual effect is each of our AI tools having on our priorities as a health system or as a community, I think, that's where we'll have to answer to the tough questions around, does it actually help? Does it actually do what it's supposed to, not simply is it accurate? So I think my hope is that the next 10 years sees more investment into AI implementation science, the science of implementing and studying does AI actually work. But my fear would be that we spend a lot of time getting better and better models, maybe even implementing some of those models, but not measuring and understanding which of those actually work and improve outcomes so that when we need to pare things back, we'll have the information we need to pair back the things that don't work so that we can shift resources to the things that do.

It also seems like it would lead to a certain amount of fatigue. If you're always throwing spaghetti at the wall, I also wonder if doctors are just like, oh my God, another initiative, another program we're trying. I have no idea if this stuff is well tested, if it's good. I suspect it's not really working that well. And I wonder if you tire out your workforce

100%. And I think we've actually looked at this a little bit. We haven't published on it, but we looked at one point and presented this at a conference where what if a system was using both a sepsis tool and a deterioration tool, don't those overlap?

If someone's getting sicker, aren't they picked up both by a sepsis model and a deterioration model because a lot of the signs and symptoms of getting sicker in the hospital are overlapped between those two? And we found that indeed, there is a degree of overlap. The timing doesn't always overlap.

One of those might fire always ahead of the other one. But the point stands that if you have a model that predicts failure of every single organ, when someone's getting sicker, all of those models, at some point, are going to fire. And at some point, that information is not going to be helpful. And there's going to be no action that you can take from that beyond that initial alert.

So I think the two things I would say are we can't just think of this as alert-based workflows. There are many, many other ways to integrate AI into health care, even predictive AI that doesn't require generating an interruptive alert and stopping someone from what they were doing. The flip side of that, though, is that we have to measure it and see, does it actually work?

Dr. Karandeep Singh is the Chief Health AI Officer at the University of California, San Diego Health. Thank you so much for being here. This was so interesting.

Thank you, Kara.

And now another installment in our occasional series "Looking into the Labs of CSAIL Researchers." The last couple of years have been all about large language models. And both academic researchers and companies around the world have raced to create better and better models, which, it turns out, has been really helpful to roboticists, even though LLMs are text based,

I'm excited about LLMs for 2 and 1/2 reasons.

That's Andreaa Bobu, an Assistant Professor at MIT in AeroAstro and CSAIL. She leads the Collaborative Learning and Autonomy Research Lab. So, what are those 2 and 1/2 reasons? Well, first--

LLMs are really good. The surprisingly good at both understanding and generating human language, unlike what we've seen before. So what this means is that I can have slang, nuanced language, or idioms, and LLMs can understand that.

Now, you don't have to have special knowledge about how to interact with an automated system by saying, for example, Alexa or Siri or Google. You just interact normally. So that's the first reason that Bobu is excited. And here's the second one.

Because they've been trained on massive amounts of human data, they now have a wealth of general knowledge and these strong, what we call them, common sense priors about various situations or scenarios in the world.

Part of tackling those scenarios is that LLMs can break things down into steps. You may have tried creating a business plan or a step-by-step approach to a project or a memo. Well, a step-by-step approach can be super helpful for a robot, too.

So for a robot, if I were to ask it something like, hey, robot, can you make me a bowl of cereal? I don't have to tell it, first, you go to the fridge, you open the fridge, you grab the milk, then you go to-- I don't need to tell it all of these steps. LLMs actually already have this prior information. So LLMs themselves can chain the sequence of commands, and they know what the robot is supposed to do. So now all that the robot needs to know or with an asterisk, but all that the robot needs to know is how to ground these instructions, like open the fridge into actions, like how do you actually open the fridge in real life.

So that's reason number 2 for optimism. And now I'm guessing you're ready for a reason number 2 and 1/2.

LLMs have a capacity to hold and maintain context across longer conversations. So what this means is that they have longer memory of what we've been conversing. So this means that we can have these multi-turn dialogue. You can maybe track what I said five minutes ago, and you have this context, this memory that kind of shapes how this conversation is going to go.

Except, she says, this ability is not as great as it could be. Take, for instance, her experience creating an image for a talk that she was slated to give.

And so I prompted it and I said, hey, generate the image with a human having a bowl of cereal in their kitchen and a laptop on the table. And the robot is serving a cup of coffee. And the LLM generated something.

And I said, oh, wait, there's no bowl of cereal. Can you please add a Bowl of cereal. And the LLM would add two bowls of cereal. And it would be like, wait, can you remove that second bowl of cereal?

And then it would remove the bowl of cereal, but it would also remove the coffee, despite the fact that I didn't say anything about it. And then I would have to say, oh, wait, don't remove the coffee, but remove that extra bowl of cereal. So there was this thing where it kind of has memory, but it also doesn't. It's definitely it's called a context window, and the context window is not as good as it could be.

Bobu has been using LLMs more in her own research, and one thing she's tried to do is to get robots to be more attuned to human emotion.

I could say something like, oh, the weather is so nice today, and because the LLM has context and it has this prior common sense knowledge, it was able to produce happy-like motion just because it knows that, oh, the weather is great today maps pretty close to the happy emotion, less close to the sad emotion. And likewise, if we prompted it with I had a tough day at work today, it produced sad behavior. So you're capable of this really nice generalization to expressions that you've never seen before without needing the human user to give any additional data. So that's why the LLM priors are so exciting.

So here we are, 2024. We have these very complicated, fairly personalized phones that we travel around with. So, when are robots going to join the party? Bobu notes that the boom in AI and machine learning is creating lots of excitement in robots, including from NVIDIA, from Tesla, companies that are both working hard on human-like robots.

What's nice about what AI has become in the recent decade is maybe in the decades prior, you would still have some level of fine tuning or some level of handcrafting. You would need to program things. And if the context changed, or the environment changed would need to change something in the programming as well. Whereas now AI and machine learning is enabling us to do a lot of these things automatically or automatize a lot of these processes. And so this gives us a lot of flexibility, which is really crucial for personal robots and for handling the varied and unexpected human environments that you would be seeing.

But-- and yes, there are at least a couple of buts here-- Robots are going to have to have a lot of autonomy in order to function in a human world.

And autonomy is currently in the current paradigms requiring a lot of human input, and nobody is going to have time to keep training the robot. And can you imagine if my robot needed as much input or as much correction as your autocorrect does. That would be quite annoying, because I am very frustrated by my autocorrect often times.

"Plus," Bobu says, "it's hard for robots to deal with the level of personalization that we might hope for." When do you want breakfast or coffee? How exactly do you want it made? Where do you want it served? And then there are thorny, thorny issues with privacy.

So right now, we're making-- whenever we use ChatGPT, Claude, Gemini, again, insert your favorite assistant, we are kind of making a deal that we're giving away our data in return for this service. But can you imagine if we are going to have personal robots in our home, they're going to collect more than just text information. They're going to collect video information about the environment because they need to be able to sense the environment. They're going to have microphone feeds, location data. And so can you-- would I be comfortable with that?

And one more small thing to consider on the road to personal robots, well, maybe not such a small thing.

Robots are very expensive. The cheapest are a few thousand dollars. Most robots are tens of thousands of dollars. There needs to be quite some hardware breakthrough to be able to bring the cost down and make it affordable for the average consumer.

So a lot of excitement, but still a few hurdles before you bring home that robotic butler.

[MUSIC PLAYING]

Thanks to Andreaa Bobu, an Assistant Professor at MIT in AeroAstro and CSAIL. And before we go here, if you want to keep up with the latest developments in computer science and AI, follow MIT CSAIL on LinkedIn. You can see amazing videos, lots of news and features, and alerts about cutting-edge research.

I'm Kara Miller. The podcast is produced by Matt Purdy with help from Andrew Zukowski and Audrey Woods. Join us again next time, and stay ahead of the curve.

[MUSIC PLAYING]