After the hype: what can AI really do for human-robot interaction?

Transcript of the Human-Robot Interaction Podcast Episode 28. Published on 9 May 2024 by Christoph Bartneck at:

https://www.human-robot-interaction.org/2024/05/09/after-the-hype-what-can-ai-really-do-for-human-robot-interaction/

Christoph Bartneck: [00:00:00] This is the Human-Robot Interaction Podcast.

I'm your host, Christoph Bartnack.

Welcome to the Human Robot Interaction Podcast. Today I have the privilege of having two esteemed guests here in the studio that I'll let them introduce themselves. Dave.

Dave: Good morning, I'm Dave Kaber. I'm here as a Visiting Erskine Fellow at the University of Canterbury. My permanent position right now is with the University of Florida. One of my main research areas of interest is human robot interaction, and I'm very glad to be here for the podcast.

Christoph Bartneck: And Julie, would you like to introduce yourself?

Julie: Sure. I'm Julie Adams and I'm the associate director of research for the collaborative robotics and intelligence systems Institute at Oregon University. My research focuses on human teaming with complex [00:01:00] systems and in the last 20 years, really robotics.

Christoph Bartneck: Thank you so much. Artificial intelligence. I'm not sure if the hype could be any bigger than this. Time for some seasoned academics to talk about Should I say this? Okay, let me rephrase this.

Dave: Seasoned is fine.

Julie: is fine. Seasoned is perfectly fine. Is that okay? I think it's a fit. Okay, okay. I should say experienced or I'm not sure. It's been 20 since we started the HRI conference,

Yes, I know. So,

Dave: That's crazy. I hadn't thought about that before.

Julie: Yeah. Next year is the 20th.

Dave: Wow.

Julie: So yeah, we're seasoned. That's

Christoph Bartneck: Okay. Okay. Let me try this again. We're going to talk about artificial intelligence. And who better to talk to about this than Julie and Dave. What is AI? Julie, what do you think?

Julie: Sure. A lot of people today think that AI is machine learning, but in fact AI is a much broader field. And we typically [00:02:00] talk about AI really representing different capabilities that humans have, whether they are through Narrow AI, which is really a focused single application of AI or or something more like general artificial intelligence, which we hear a lot about in the media today, which is intended to represent humans capabilities. But we're still pretty far away from that. And we'll talk about that more in a moment.

Christoph Bartneck: Dave, what do you think?

Dave: Yeah, I think my perspective is fairly broad. In industrial engineering for many, many years, we focus on decision support systems, which is basically using, even spreadsheet based tools sometimes computer languages for programming algorithms to provide a decision basis for human decision makers.

AI in my view is Any type of tool that can help a human being make a decision and oftentimes the technology is inspired by some neurological basis. Largely the tools that we've been working with for many years have been pattern matching

Julie: I [00:03:00] think that's key because a lot of times AI is mimicking human capabilities and it's sensing, actuation, things of that nature.

Christoph Bartneck: All right. Are there different types of AI?

Julie: Yes, there are. So there's narrow AI, which really is a system that uses AI to perform a very specific task. Something like helping with making decisions or helping to process sensor information to identify pieces of information and say, image processing. And then there's artificial general intelligence, which is intended to be more representative of human capabilities. So how we reason, how we. think, reasoning, cetera. And we still haven't achieved that yet.

Christoph Bartneck: Dave, what are your thoughts?

Dave: I think Julie covered some important categories. There's AI, there's AGI, now there's GAI generative artificial intelligence. The technologies purport to do different things. I would tend to agree that I think we're still very [00:04:00] much focused or limited to narrow AI as you've characterized it.

I think there are some attempts currently to develop systems that can generate novel output based on training to a certain set of inputs. Yeah. So there's several different categories. I agree. I really don't think that we have any artificial general intelligence at this stage. And I'm also a bit dubious of the generative artificial intelligence being anything more than really complex lookup tables, frankly.

Julie: And I think it's really important that people recognize that AI is not new. It's existed for Over a hundred years in various forms and narrow AI is something that we encounter every day. If you do a Google search, for example, that's an example of basic artificial intelligence algorithms.

Autopilots and aircraft have been around since 1914 and this is something that is also a variant of narrow AI. So I think it's important people recognize that it's not a new technology. The generative AI is a newer variant of [00:05:00] artificial intelligence.

Christoph Bartneck: How has AI been used in human robot interactions specifically?

Julie: I think a key thing is that we've been using it for years. I mean, my research has been going on now for 35 years, and we're using artificial intelligence at the very beginning, just to even understand what the robot could perceive in the environment. Robots doing path planning. From the early 1990s is another example of basic artificial intelligence that's existed.

And a lot of these systems, even today, are still semi autonomous. While you hear a lot about autonomous vehicles. If you look at robots going into uncertain and dynamic environments, We aren't going to be talking about fully autonomous, highly intelligent robots for a long time. You're still going to have to have a human involved in the interaction with the robot for the system to be effective.

Dave: Yeah, so I think Julie [00:06:00] hits the nail on the head. I think AI has been developed for functional purposes. For me, it's hard to separate the basic functional purposes, for computer science applications from the extensions to human robot interaction. So for example, yes, path planning is, I think, a very good example.

I think that image processing is also an important example. To me, that fundamentally comes from computer science, but there might be HRI applications of image processing. I think of industrial robots for doing recycling, sorting and processing waste that use incredible image processing, very fast image processing algorithms.

In my view, there's also more advanced forms where we've actually used machine learning tools for attempt to provide robots with some capability to sense like the states of a human being and to use that information for collaborative task performance. So in my view, that's a little bit more advanced but I think all the applications are very functional oriented and they can be extended to HRI. But fundamentally, I think that they're developed for, basic CS computer science [00:07:00] purposes.

Julie: I think that there's, a lot of basic artificial intelligence, especially some of the machine learning capabilities, as Dave said can be used within robotics. And, oftentimes you're talking about more traditional kind of rule based systems or other types of algorithms that get applied. In our own research, we've done something similar to what Dave talked about, which I think is more kind of the more recent cutting edge work where you're taking things like physiological sensor data, processing it, and using machine learning to identify the human's workload state so that you can adaptations either in the interaction or the autonomous allocation of tasks between team members.

Christoph Bartneck: So The word machine learning has been thrown around here quite a bit already. When we talk about learning, that implies that somehow the [00:08:00] machine learns or gets trained. How does that work in human robot interaction?

Dave: So my students and I have used machine learning for training different algorithms on human emotional states or for human emotional state recognition. So the training requires that we have a very large database of historical performance or historical responses from a human being. That data has to be processed, cleaned statistically aggregated. Sometimes we do signal processing to extract particular features. Those features are used as input variables into the machine learning algorithm.

Basically the algorithm is, in my view, it's like a system identification tool. So, in other words, we have a set of inputs, and those inputs are connected with a finite set of outputs that are defined in advance. So we use some gold standard for defining the output. The machine learning algorithm basically establishes all these interconnections between the inputs and outputs.

We train it on a fraction of the data that we have historically or that we collect through experimentation. And we also use a portion of the data for [00:09:00] testing. And sometimes the learning can actually be statistical based learning. In an earlier exchange, Julie referred to this.

The inputs may be stochastic in nature, the outputs can be stochastic in nature. In that case, we would use distributions to represent the inputs or the outputs. But nonetheless, I mean, the key thing is the algorithm remains a pattern recognition or a pattern matching algorithm. That there really isn't causality there, or the algorithm does not have the capability for explaining why those relationships exist between the inputs and Maybe that's more than what you're looking for right now, but, yeah.

Julie: No, I a great definition or example. And I think one of the key things if you look at human interaction with these systems is how do you get that data set that Dave was talking about, and I think we can talk more about that as we go forward.

Christoph Bartneck: This is roughly what I was getting at. When we want to learn from humans, have to observe humans. Which means that [00:10:00] you cannot just scrape the internet for billions of data points. Right. So, what are the constraints that we have when we learn from humans?

Julie: In our work, anybody who's trying to do human systems really wants to have what we call ecologically valid data collection. Meaning that the ways that we're collecting the data represents the context in which the tasks will be done in the future.

We really are trying to develop these very advanced user evaluations in order to collect that data. But anybody who's done user evaluations know that you're somewhat limited in the number of participants you can have, the amount of data that you can collect. We're running evaluations that are 70 minutes long for just the data collection.

And that's exceptionally long for most people. studies. You're going to have a very good data set, but humans are highly variable. They never do exactly what we expect them to do. And so you still have a lot of noise in that data, as Dave alluded to, with having to clean the data set, [00:11:00] etc.

And then you also have a limited amount of information Uh, that you can use for the machine learning. So we recently ran an evaluation with 54 participants that took us nine months. And we still have to be careful about how we use that data when we're training our models, not to have overfitting and some of the other issues that you run into. Dave, you're smirking. Yeah, yeah, no, I'm good. What were you going to say?

Dave: Well, so, uh, I mean, now we've moved into deep learning neural networks. And in my view basically you have an architecture there that's unconstrained. When we first started working with, like, feedforward neural networks, We were always concerned about developing parsimonious algorithms, where we were carefully managing the hidden layers and the number of nodes, and so we could manage the patterns, the number of patterns that are identified between the inputs and the outputs.

But now, with the LLMs, it's just the wild west, frankly. You can have hundreds of layers, hidden layers, you can have hundreds of nodes within each layer. And so, you just don't even know how the, quite [00:12:00] frankly, it's very difficult to determine how the algorithm is establishing the interconnections between the inputs and the outputs. And, if you're using something like that for target recognition and classification, and the Air Force is going to deploy a weapon on the basis of the output from that network, then, there really needs to be some explainability of how the algorithm works. And, I think constraining the architecture is a critical. dimension and you can really get overfitting to data when you have like unconstrained architecture.

Julie: And another important thing to realize, as Dave alluded to earlier the outcomes of these models are really dependent upon the input data. So, in our case, where we're working in wildland fire response as one of our examples are disaster response. We have the ability to potentially use our first responders to train with their robots just like they would train with humans to collect an initial data set. However, every first responder typically will tell you that Every situation is different.

So that means that you're going to have [00:13:00] tasks that they have to do that the machine learning algorithms have never seen. So then how do you use those machine learning algorithms if you aren't able to learn in real time, which in a lot of cases you're not going to be able to. How do you use those algorithms to handle those situations that haven't been seen before and not come up with incorrect classifications or detrimental classifications for the robot's response?

Christoph Bartneck: So one step forward or let's say a difference that we see to the currently very popular large language models is the training data. Whereas large language models have been able to utilize millions and billions of text and images and can build upon this. Is something like this realistic for HRI?

Julie: So, I think there are pieces of HRI that can benefit from large language models, but I think it's going to be limited. In my case, where [00:14:00] we are working in these highly uncertain and dynamic environments, Our inputs to our systems are not text from the Internet. They are not natural language coming out of an individual that's been recorded or image data. You're going to be limited on what you can learn and how you can use that information.

Additionally, the domains I work are very specialized. So you just don't have the data set available to do effective large language models. If you're in a medical domain, you have to have appropriate medical, um, datasets. If you are in wildland fire response, you have to have appropriate datasets, and you aren't going to have lots of images that are appropriate for a first responder or a wildland firefighter in what you see in the media and the text and information you have from the media.

Christoph Bartneck: Dave, you previously called this large language models a large lookup table. Could you maybe elaborate on that?

Dave: I'm hopeful that they will be useful [00:15:00] for for HRI. Because I think that many of the automation technology interfaces that we have. Today are quite cryptic and they create usability or usefulness issues for for operators.

I think a fantastic application would be if, you know, if we can take output from an automated system and we can pump that into an LLM and that we end up with some natural language output that's more understandable for the user. That's a whole nother set of problems, because I think that there has to be conciseness in communication.

It's got to be pointed. What I meant by lookup table is, I mean, fundamentally for these models, I think the core technology is it's a deep learning neural network. Basically the network is being trained, like Julie said, to a very, very large, data set. Often the source is unspecified or very, very broad, but there are interconnections that are being trained within the network and that's based on a training data set.

And that's what the algorithm knows. Factually speaking, it does not have the capability to generate new interconnections unless there's additional data that's used for training. I think Julie [00:16:00] brings up some, alludes to some very important points. Important practical limitations.

I'm hopeful that LLMs can be useful, but that they're very large, right? So the models, well, whatever large is today, but they're large such that oftentimes you do need to access the cloud to take advantage of them. Julie's talking about some field applications that she has a lot of experience in, and, I don't know how you could access the cloud for some of those, like for wilderness rescue or, and then furthermore, if you could load the LLM locally onto the robot, I don't know what that would mean from a computational perspective, just based on the technology that we have today. So I think in terms of fielding, there's a lot of practical limitations at this point.

Julie: That's an excellent point. And one of the points I was going to bring up because there is often especially within the research domain. A lot of assumptions about fact that communications are available and that you can always access the cloud or that you have enough computational capability.

And frequently, the systems that you want to take into these environments are not going to have the computational ability [00:17:00] to do these large language models, especially on the fly. And you really need those computational resources and your battery, your power consumption payload to be dedicated to the task at hand.

I do want to go back just for a second about how do we generate data though, right, because we talked a little bit about the human inputs, but there's a lot of work being done right now in what's called Sim2Real or Simulation2Real, where people are using simulation to try to generate data for training these models especially for deep learning, not necessarily for large language models.

And this a great way for generating data for some systems, but I would argue that for these highly uncertain dynamic environments they aren't going to be sufficient, especially if you're talking about multi robot systems or heterogeneous robot systems going out into these domains. And that's because the simulations Can only represent a certain level of the physics and the dynamics and the uncertainty in the [00:18:00] environment, the robotics capabilities in order to be able to train them.

We're seeing great success, for example, training legged robots to Navigate complex environments, you know, it's a great way to do that or to do some manipulation training, but I think emulating humans in the sim to real becomes quite complex because I don't know that we have the ability to emulate all aspects of what humans will do and Again, trying to get the complexities of the dynamics of the physics in the environment

Music

Christoph Bartneck: So what can they do for HRI?

Dave: We've been talking about some important applications. I think that there's a whole host of other applications for which LLMs could be useful. I think it depends on how you define robotics, frankly. I'm thinking of soft bots. So in the financial [00:19:00] industry these kind of models can be quite important.

Some of my students who were previously working on human emotional state classification for robot situation awareness, now they're taking those same models and they're applying them to financial data, for forecasting market trends and stock pricing and that kind of thing.

So I don't know that we touched on that, but that's a lot of what these models do and. Yes, prediction of human behavior in a wilderness rescue situation is important, but what I'm saying is that there's a broader spectrum of applica like financial applications, medical applications where I do think that these models would be accessible and they can serve as a very effective intermediary between a form of automation and a human being to explain what the state of the system is so that the human has greater situation awareness and can make better decisions.

What I hope is that LLMs can actually be used for AI to say hey, I'm only 60 percent certain about this prediction and you should be aware of that and using it as a basis for decision making. Or, you know what, I don't think I've got this right. I could use a little help here. You know, could you intervene, right? We need [00:20:00] humble AI and I'm hopeful that LLMs could be like an intermediary to achieve that,

Julie: I agree with that. I think there are actually some great HRI domains for this kind of technology. If you look at potential robot systems in the home to help with adult care if you look at manufacturing these are controlled environments.

In the manufacturing domain, you have the opportunity to really collect the data you need to train the large language model, which is very different, I think, than some of the ways these models are being trained today. But I think that's feasible. In the home a lot of the things that the robot would see from a vision perspective are things that can trained from the internet.

I think dialogue wth an individual. Either from an emotional perspective or a fundamental perspective of it's time for your medication, for example can be. Enabled as well as enhanced with large language models. So I think these are great domains where you can start talking about bringing this technology [00:21:00] to enable the capabilities.

Christoph Bartneck: Sometimes it's interesting not to ask what it can do, but also what can it not do. And And of course, right now, since we're in the AI hype, everything seems to be possible with large language models. And your point of view. What do you think large language models cannot do?

Julie: So I'm a skeptic of the technology, even though I've worked in AI for so long. I think that there are a lot of limitations on its ability to generalize. In particular, I think, aside from the limitations of getting the data sets and training the systems, are means for it to facilitate interaction with autonomous systems, but it's not going to be representative of how we think as humans.

It's not going to be as smart as we are. It's perceived in many ways inaccurately as [00:22:00] having these capabilities. Especially by individuals who don't necessarily understand the underlying systems and how the machine learning works. So I think there's a danger in people really believing these systems and taking them as truth over trust in these systems.

Dave: The only thing I'd add is that I don't know that there has been sufficient research or structured research on What would be the most effective form of LLM use for human collaboration with robots? So my concern is that we may have a technology that enables a robot to better understand the state of the human, and then there's some attempt to use the LLM to provide information.

And so I don't think we know what the nature of the information what type of information aiding should occur, what the frequency of the aiding should be what the modality of presentation . So I'm worried about how the LLM communicates with a human, and I don't necessarily know that that it's going to solve the human autonomy teaming communication problems that we have.

The fact that we have automation that really cannot operate as [00:23:00] like a true partner to a human being right now. I don't see LLMs being like an immediate solution to that. I think there needs to be more structured research to determine how they can be effectively used to solve human robot interaction problems.

Julie: I think a great example of that is what we've seen with autonomous vehicles. In that autonomous vehicles have all this training data, but yet they still aren't able to drive safely in a lot of context. And I think there's also a privacy issue as well. Because in order for these systems to be used in real world environments, they're going to have to continue to learn and be adaptive.

And that's not something that I think most of the research is focused on. And in order to do that, they have to continually collect data from the humans that they're interacting with. So that then also becomes, raises a number of privacy issues as well.

Dave: Returning to the hype earlier when we were talking about applications and uses of AI I don't think we got into some of the really fantastic applications that, are being presented. I think there are [00:24:00] some researchers that have concepts of AI being collaborative partners to human beings in most circumstances in the future. What I want to say is that there's a lot of fantastic applications that have been identified, and that for persons who might not be, closely familiar with the technology, that those applications are, in my view, are real stretches of where we're at currently in terms of AI that's been developed.

Yes. And to Julie's point, there are limitations there that can really lead to very serious issues, problems and system operation. And, that in my view, that's like a business proposition or perspective driving, or pushing the technological development. When from an engineering perspective, I don't know that we're actually there or that it's totally feasible.

Christoph Bartneck: Let's drill into that because let's take Tesla. We talk about the Optimus humanoid robot that's being developed and it's just one of many humanoid robots are currently being developed. Yes. [00:25:00] What do you make of it?

Julie: I have mixed feelings on humanoid robots. I tend to focus on, ground based robots or marine robots or aerial robots and not humanoids. I think there are Circumstances in which you do need a robot in the form of humans especially if you're trying to use robots in a environment that was already built for humans. Space Station is a great example of why NASA developed Robonaut. To be a humanoid form because the Space Station was built for humans and not necessarily robots. And I think there's a number of circumstances like that. Our homes are a good example, but do you actually need a humanoid robot in my perspective is not really in most circumstances.

Dave: I would agree and I think, so the cases that you identified, like aerial robots, ground robots, in my view, those are more narrow forms of AI that are spoken, focused on very specific functions and applications.

Whereas when we talk about humanoid [00:26:00] robots like Atlas, for example, Oh. Who knows what the range of applications is, right? And it comes off as like AGI, like artificial general intelligence. And I think to Julie's point, it's, let's see how we can do with narrow AI for aerial applications or for ground based applications before we start getting into, creating general humanoid robots that have AGI that are performing a whole host of tasks. I don't really think, personally, I don't think we're there yet.

Julie: I think we're a long ways away from it. All need to do is go back again to the autonomous vehicles, right? We still have so many limitations with those vehicles however many petabytes, zegabytes, whatever it is that they have of data.

Humanoids are going to need Even more data, because they've got more functionality more degrees of freedom, if you will, in their capabilities and in order to be able to actually be effective, I just think we're very far away for. I also fear that for the [00:27:00] general public, there is going to be a perception of these robots having human equivalence when in fact they don't have that, right? So it's a misunderstanding, an incorrect mental model, if you will, of how these robots actually function. That is going to lead to, I think, a lot of concerning situations in the future.

Christoph Bartneck: But when we look at this example again of Optimus or Agile Robotics, who claims that they will enter mass production of the humanoid robots, BMW making a claim they're going to buy in to use in manufacturing. Optimus, supposedly going into mass production soon, or it will solve all kinds of things. There seems to be a lot of promises and developments right now in this area. And I wonder, and I'm less expert on this than you are, know can they actually be useful [00:28:00] for us?

Julie: I do think so. Agility Robotics was started by Jonathan Hurst, one of our faculty members. And the Digit robot is a robot that's being tested in Amazon warehouses right now. And I do think that there are contacts for that robot.

However, I will say that agility is not presenting their robot. It's a humanoid based robot. They are not presenting their robot as having AGI. They're not presenting their robot as having this very broad application. They are designing it for circumstances that I think are well defined and working in this warehouse environment to try to alleviate some of the repetitive types of injuries that humans can get doing that work as well as trying to help the humans be more productive in their tasks. And so I do think that there are circumstances there where you aren't going to have. As human, a looking robot, even though it's a humanoid that people can use [00:29:00] effectively.

Dave: I agree. I think that they can be helpful for certain applications, but I think that we need to define the problem space. We need to constrain it. Otherwise, I think the human interpretation can go like way beyond what the capabilities are. So you were going after like the labeling that we assign, right? So autopilot. So I think that labeling for customers can lead to a miscalibration of trust in the system. For the everyday user, the perception that, the vehicle can totally drive itself. Because right now what happens is that you have users that trust it way beyond what they should.

Christoph Bartneck: But since all these claims were made about mass production of humanoids, it seems like, is this just hype in terms of marketing? Or you mentioned that the Optimus robot being used in showrooms. Is it just media hype right now or are we really at the brink of, well, next year we're gonna have mass produced humanoids they're gonna be in every factory, it's gonna be [00:30:00] everywhere and do work for us.

Julie: I'm skeptical that will be the case. I think we're probably still, I think five years is optimistic. Probably more like 10 and except for these very well defined scenarios, that I would argue are not general public. I, for example, just came from South Korea. I was there for 12 hours and in 12 hours I saw two robots. One in the airport that was I would call it more of a research humanoid form, meaning a mobile base with more of a facial expression type of system. And then a lawnmower robot. I think the true humanoids, the ones that, Look like humans with legs and arms and expressive faces. I think we're still probably at least 10 years away from having a really More general system that could be deployed in the general public probably longer. Even then I think it will still be narrow AI. It's definitely not going to [00:31:00] be a general AI system.

Christoph Bartneck: We've been in academia for long enough to see these hype cycles. And currently we, I guess, agree that there is a hype. What will happen when this hype disappears? What impact will it have?

Julie: I worry that we are going back to the A. I. winter of the eighties that I didn't experience, but I've heard about. I think that's probably true for all three of us that we heard about it. We didn't experience it.

Dave: But you're right. There was a winter and now this is really a resurgence in my opinion.

Julie: And there's definitely a lot of companies overselling their capabilities. And it's not just the big companies, it's the startups. It's, we see this repeatedly, whether it's robotics or AI or any type of technology. There's a lot of overselling of their capabilities, as you've alluded to, because fundamentally it's, the business case for the company to [00:32:00] exist.

And how do they, in a startup case, raise more money? And if they're a publicly traded company, how do they, make their stockholders happy? I do fear that something like the AI winter is going to occur, and I would predict in the, next 24 months to 48 months that will happen, maybe even sooner.

Um, and, uh. That's a juicy nugget for podcast.

Wow.

I do think that that's the case, and I feel very positive about that. That's my business side coming out.

Dave: So my view is, research takes time. And there are. Some NSF programs right now, like M3X, that I think that they are. Attempting to fund more research on how humans interact collaboratively with autonomous or semi autonomous systems.

And for many, many years, human factors has been in the role of cleaning up after the parade. Yes. And so Julie's correct that, there's technological advancement. There's a business proposition that's driving that. [00:33:00] And then because of The timeline for the research process then we kind of follow behind and, then provide design principles and make sure that we get it right going forward. I think that the hype's going to die down. Then there'll be systematic research to identify, establish theories and principles for how we're going to effectively design. AGI or GAI going forward so that it's useful in these types of applications. Yeah.

Julie: There's what I call important tipping points for technology development. And I think in robotics, a really important Milestone was ironically the iPhone because it miniaturized computing and miniaturized sensors and really enabled a lot of the robots we have today.

The robots I use for my doctoral dissertation, you could not take home. They weighed too much, right? Now I can put a robot in my pocket. And I think that was really critical and I think with the technology development for large language models and some of the AI that we're seeing today, I think it's an important tipping point.

But I think it's just a milestone in the step [00:34:00] towards achieving better systems in the future. Yeah. But I worry, with the AI winter that I'm predicting, that will also set us back, right? I worry that it's going to cause a withdrawal of startup funding, a withdrawal of funding in companies, and a withdrawal of funding at the federal level for researchers and that would even set us back further.

Christoph Bartneck: Do you think that large language models are our way towards a general AI?

Dave: We had some exchange about this previously. My concern is that until we come to AI that can actually do causal based reasoning. And, there's a whole community on causal reasoning AI. And so there is research in this direction. But if we think about the important applications that we'd like to target, like medical applications there has to be accountability and, of course we want to have a great degree of reliability and so I think that that there is a need for us to develop technologies that actually have the capability to project two [00:35:00] way and three way interactions, which are difficult for human beings to do, period. But the machine has to be able to do that, and in a novel way, too. And there may be a set of circumstances that it hasn't been exposed to previously.

And not, it has to exhibit knowledge based performance, right? So use some mental model of a set of relationships and use that for solving a problem. I think until we get there, I don't see LLMs as being the key to AGI, frankly.

Julie: I think a key aspect of humans is that we are highly adaptable, right? We can encounter situations that we haven't seen before, and we can take the knowledge that we have and adapt usually to be effective of dealing with that situation. Now with artificial intelligence they can adapt, but we don't necessarily know what the outcome is going to be and one of the biggest One not the biggest but a big research question right now is how do you verify and validate what's going to come out of these systems? And then you talk about certain domains like [00:36:00] medicine. There's going to be regulatory needs. And in order to get regulatory approval for some of these technologies to be put into the medical domain, you're going to have the verification and validation capabilities of this technology.

And we already know, like when we try to verify and validate Even narrow AI for concerning circumstances like medical that the response of the system is often dependent on the human's reaction or the human's inputs. So, I think we're going to be very challenged to develop the regulatory compliance capabilities. The verification and validation of the systems, but also allow these systems to adapt. And we're still very, very far away from that.

Christoph Bartneck: So if we take the example of the city of New York, which had a chatbot advise businesses, and apparently it didn't go so well, is that an example of [00:37:00] problems that we might face with large language models?

Julie: I think it's a great example. And, having worked in industry and human factors, human acceptance of the technology is really important. So you don't want to develop a poor system and put it out there because then that leads to a negative reaction by the potential users of the system. You want to have a system that is going to act as anticipated and be easy to use and be accepted, but also not overtrusted and be helpful and useful.

Dave: The only thing I'd add is that I think Google's on the right track conceptually speaking. So their AI team is developing different models for different domains of knowledge or different functions. And then the idea is that these models would be put together to create some form of general intelligence.

And so it's, like from the human perspective, when we talk about cognitive science, there, we refer to different memory structures, and so [00:38:00] mental models is one, one conceptual structure. And so the idea is that You have an LLM or Deep Learning Neural Network that represents different mental models, and these are all concatenated or combined together to create an AGI, and then there's some communication among those mental models that would solve, that would address unique problem situations, or that there's The intelligence, the artificial intelligence would be able to address novel circumstances.

So I think that, like that layering and this integrating multiple models together, conceptually speaking, that seems like it's on the right track.

Julie: I agree. But I think we also have to be careful if you think back to expert systems. So developing these individual domain models, we have to be careful that we don't end up with systems that are too brittle, like the expert systems were.

And then I think if we go back and we look at some of the early work on the subsumption architecture by Rodney Brooks, which was the idea of build these individual very specific capabilities and then you bring them together to create more complex behaviors. [00:39:00] There were a lot of emergent behaviors that were unanticipated and that could also be a consequence of doing these individual model developments and then bringing them together, then end up with outputs from these systems that are unintended and perhaps detrimental.

Christoph Bartneck: So the future of AI is going to remain interesting, and we're going to be around to observe some of it, I suppose.

Julie: Hopefully we're going to be observing quite a bit of it.

Dave: We're so seasoned,

Christoph Bartneck: Okay, very last question. Do you think general artificial intelligence will arrive before you retire?

Julie: No.

Dave: I don't think so, but I do think that my daughter is going to see a lot of AI development during the course of her life and I hear the projections from the Microsoft team, and I think that they're just fantastic.

They sound, really great. Like having a humanoid AI robotic partner that goes with you [00:40:00] everywhere and that can support you in different types of tasks. Has a really tremendous knowledge base. I can't imagine, having Like a friend like that if you will in the future, and it's exciting It sounds exciting and for some people even sounds a bit scary, right?

And so I think it's important to do things like this, we're trying to get information out there to help people understand what the state of the technology actually is and how it can be used effectively. So no, I don't think it's going to happen in my lifetime, but I'm hopeful for would follow after

Julie: Yeah. I agree. I know when I plan to retire and it won't be before then but I do think that, the generations that are, starting grade school at this point are going to see tremendous change in their lifetime, just as we've seen, right? If you think back to where things were when we were that, young children, technology has advanced dramatically. And I think that will continue to speed up over time.

Christoph Bartneck: Then let's hope for the best. Julie, thank you so much. Dave, thank you so much for coming.

Dave: to do it. Thank you.

Julie: It's been a pleasure to have you on the podcast.

Dave: Thank you, Christoph.

[00:41:00] I'm not sure what I'm doing here.