Carnegie Mellon University
Software Engineering Institute

SEI Cyber Talk (Episode 8)

Artificial Intelligence and Machine Learning – Hype vs Reality
by Rotem Guttman and April Galyardt

Rotem Guttman: Hi, I'm Rotem Guttman.

April Galyardt: And I'm April Galyardt.

Rotem Guttman: So April, I've been meaning to ask you a question. So if I believe the advertisements and the marketing materials that I've seen, artificial intelligence is going to free us all from having to do any work, and we'll all be able to retired and relaxed and enjoying our lives. If I believe Hollywood and popular culture, it's going to bring about the apocalypse. So my question for you is: Should I be stocking up on suntan lotion for a tropical vacation, or canned goods for my bunker?

April Galyardt: Both can't hurt. So, the answer is-- I mean, with any new technology there's always a huge amount of hype, and never believe the hype. Do we remember Ginger? That's changed almost nothing except how airport security gets around. So AI has already changed more than that, but-- or machine learning-- but it doesn't work nearly as well as the hype.

Rotem Guttman: You mentioned just now machine learning, and so one of the things that I've seen the lines get blurred a lot, especially in the popular culture coverage, is AI versus machine learning, and even what people mean when they say AI tends to be very vague and nebulous.

April Galyardt: That's actually something we've been struggling with a lot lately, and trying to communicate to different stakeholders and congressmen, that there is a difference, that the difference matters. So in general, AI is-- it kind of comes from the 1950s, just anything that might have anything remotely to do with the eventual creation of an artificial intelligence, which includes statistics and computer science and robotics and all kinds of things, whereas machine learning-- right now the language very much refers to the intersection of computer science and statistics in terms of learning from data. So from the big picture view of AI, machine learning is part of that complex.

Rotem Guttman: But the AI solutions that I've seen marketed as AI solutions as far as I can tell are usually narrow AI, AI that's focused on one particular task and really isn't going to generalize for-- yes, it can tell a cat from a dog, but it's not going to drive a car. And so I think that one of the things that I've been seeing a lot of coverage about is all of these great advancements, but not a lot of coverage about when they can actually be used. So what makes a good candidate for using AI techniques?

April Galyardt: So there's a couple of things, and you actually hit on one of them. One of the biggest things right now-- and these kinds of things are much more in the area of machine learning, where you're doing the same kind of thing over and over and over again. Those are the kinds of things where machine learning-- there's lots of data and it can pull from it and give you something that's really informative. So if you think about Google and searching-- lots of people search, lots of people click, and so you can put that into a machine learning algorithm and that feeds information about the next click or the next search. Same with image categorization. We've got lots of images, they need lots of labels, and so labeling an image is something that has to be done over and over and over again, and once you teach the machine how to do it, it's faster than humans.

Rotem Guttman: But only if the machine is doing the same thing that it was trained on.

April Galyardt: Yes.

Rotem Guttman: I think one of the pitfalls that we've both seen over and over is where the way that the solution has been trained or developed doesn't exactly match the way it's being used; you can have some surprising behavior from the system.

April Galyardt: Yes, that is in fact one of the biggest problems. It really does have to be the same task. And so if you're training on a particular set of images that had been taken at a particular resolution and suddenly you switch to images that were taken with a different camera and a different resolution, then suddenly your algorithm doesn't work anymore, and that's just the simplest example of context really has to matter. We can get into stories-- the Google flu trends is one of my favorite examples of how things go wrong when context changes. So Google had flu trends, and they were trying to use how people searched for different types of flu symptoms to estimate how bad this year's flu season was, and for a couple of years it worked beautifully, it was amazing, they were making accurate predictions faster than the CDC, the Center for Disease Control, and then one year it all blew up and they overestimated how bad the flu was by more than a hundred percent. It was just way, way off. Because what had happened, they had used the results from their flu trends to change the behavior in terms of how people were searching and clicking on different flu symptoms, and so now the behavior has changed and now the model that had been trained, it's not exactly the same context anymore and the model made very bad predictions.

Rotem Guttman: So it was a technically unrelated change that changed what it was measuring so that now-- and it obviously has no way of being aware of those external changes.

April Galyardt: The model has no idea.

Rotem Guttman: Yeah. I think the most amusing case I've seen was actually not that the parameters changes but where the system-- depending on how you design your AI system, it may or may not be prescriptive about what data you're actually giving it. And so in a system where they were allowing it to essentially train itself and learn what the criteria were, they allowed the system to play Tetris, a very simple game, and it watched a bunch of people playing Tetris and it saw-- and the information that was basically given was these videos of people playing them and the inputs. It trained itself, and what it did is it played until the board filled up, and then right before it lost the game it paused and just left.

April Galyardt: You can't lose if you--

Rotem Guttman: "Can't lose if I don't unpause the game." They're like, "Well yes, but that's not what I meant," and that's something that we see in more serious use cases of AI over and over and over again, where yes, technically the system is doing what it was asked to do, but it's going to optimize exactly what you asked it to optimize, not what you meant to ask it to optimize.

April Galyardt: Yes.

Rotem Guttman: And those can lead to some very interesting edge cases. Now, what do you think the most important use case of AI outsource going to be in the next five, ten years? Where do you see it really changing things?

April Galyardt: Important-- that's a hard one. There's a couple of easy, low-hanging fruit right now that I'm seeing. You were just talking about kind of reinforcement learning with the Tetris. So one of the places that I think that's going to get pushed into first is kind of optimization scheduling and logistics and that sort of thing, and that's going to change a few things. In terms of the most important, it's hard to even anticipate because most of us can't even drive around without our maps open now. Like people new to Pittsburgh, it takes-- Pittsburgh roads are very interesting.

Rotem Guttman: Five hundred feet, four lanes on a bridge? Sure, we'll do that.

April Galyardt: Right, or if you're looking at a map, everybody has this experience where you're driving along and you think, "Where's my turn? Oh, the road's up there." And so everybody around here is using the maps and can't get anywhere without them. So I think a lot of little changes like that are going to add up to more big change.

Rotem Guttman: One of the things you mentioned for example is logistics. So these systems, they don't always make the perfect prediction, and so what do we do about the case where the system says, "Turn right here," and the driver who's been doing this route for 20 years says, "No, I know this route. It's faster for me to go straight." How do we solve that trustworthiness problem?

April Galyardt: Some of that comes with experience. So I've had that experience with navigation in my car, and it's going, "Turn right here," and I'm going, "No, I want to go this way," and then I go that way, and it's like, "Oh, there was an accident there that you didn't tell me about. You just said 'Turn right.' I see." And so now it's going to increase-- I went against it and I made a bad decision, and that's going to increase the likelihood that I'm going to do what it tells me the next time.

Rotem Guttman: But I want to touch on a point that you said there, is it didn't tell you there was an accident there. Now, if the system had told you, "Turn right, there's an accident on your usual route," you would have been far more likely to listen to it in the first place.

April Galyardt: Exactly.

Rotem Guttman: And so one of the things that we've actually been seeing a lot in these AI systems is actually designing your system so that it can explain, at least in a cursory manner, why it's made the choice that it made. It's a lot more likely to be trusted.

April Galyardt: Yes, although there's trickery there too. So for example, there's lots of examples coming out of psychology where if you can present two scenarios that are basically equivalent and you word it one way, you can make 75 percent of people choose this option, and if you word it a different way, you can make 75 percent of people choose another option, and so it's not just the explaining but it's also how you word it and how you present it, because you can still push people one way or another.

Rotem Guttman: Mm-hmm. So how you present it, that's another very important point, is a lot of times we see these AI systems making what are-- I always call them recommendations regardless of how they phrase it-- making judgments, so to speak.

April Galyardt: Yes. I like the "recommendations" word.

Rotem Guttman: Yeah. But one of the things that I very, very rarely see presented along with those recommendations is a confidence measure, and that's something that-- I don't think I've ever-- have you seen an AI system, any way of designing an AI system, that doesn't give some sort of uncertainty. We have that available usually.

April Galyardt: I have seen a few of the better systems start to present that information. But again, presenting the uncertainty, there's still a problem of interpretation. People tend to be very weird and personal about how they interpret probabilities. So you can think of a weather report. And so if it says there's a 30 percent chance of rain, what does that mean? And that can actually vary depending on context. If you're in Central Texas and the weatherman says there's a 30 percent chance of rain, that means it's not going to rain; leave your umbrella. If you're in Pittsburgh and there's a 30 percent chance of rain, it's going to rain 30 percent of the day. And so those interpretations of probabilities and the uncertainty measures-- that's something that we're going to have to put a lot more thought into how do we present and how do we help people interpret that uncertainty correctly, because if you look at psychology and a lot of the cog-sci stuff, that's something that people are not intuitively good at and they need a lot of training.

Rotem Guttman: Absolutely. One of the things I think I've found most useful is actually using visualizations in order to show them. So instead of showing them, "For Texas and for Pittsburgh it's 30 percent. Good luck interpreting that," actually being able to pop up something on the screen when you're giving the weather report that shows, "Okay, here's, over the day, what the likelihood of rain is." And so if Texas, you see one low line, and then in Pittsburgh you see, "Okay, here's the two times during the day where it's going to be a complete downpour," which, if anybody comes to visit Pittsburgh, wait ten minutes; the rain will usually pass. But it is a case where people are able to ingest things visually, I've noticed, a lot better than they can just by giving them the data in a raw form.

April Galyardt: Yes, that's definitely true, although one of the things that I've seen-- like some of the interfaces I've seen for machine learning platforms, where they're giving users input, there's a prediction or a risk score, and they're presenting the same risk score three different ways. It's, "Here's the numbers and here's the bar chart and here's this," and it's all the same number, and there's no extra information there. That's not helpful.

Rotem Guttman: Absolutely. So I think making it clear also what the data that you're presenting is is really difficult. It can be challenging because you don't understand in what context the person is kind of ingesting that data a lot of the times. And so it's that misapplication of that data that I've seen over and over, where people say, "Oh, the system said this, and so I did this thing and it didn't work," and it's like, "Well no, that's not what the system said." And again, I put that still on the system designer because you didn't convey your message well enough, and I think a lot of the-- there's certainly no shortage of challenges in artificial intelligence at the moment, but I think a lot of the areas where we're not making enough progress is at the interface from where the system ends and the human begins and how we actually interact between those two.

April Galyardt: I think that's absolutely correct. I think a lot of the people who've been designing the machine learning algorithms, they-- it's called the expert blind spot, where you've spent so long learning all the things and you're an expert and you forget what you didn't know when you started, and so they aren't necessarily very good at communicating all the little things that are necessary to interpret, and I think we're going to need to come back and fill in a lot of those gaps and present things. If we go back to the weather report, you've got your prediction, but then you're like, "Okay, that's a weird prediction," and you can click on the radar map and, "Oh, there's a storm. It's coming. I don't know when it's going to get here." And so that what extra information needs to be available, do we need to be able to provide, and that's going to be highly context-dependent. I mean, if we're talking about recidivism predictions for judges and parole boards, that's a very different thing than if we're talking about predictions for what we're going to search next.

Rotem Guttman: So for some of the context that you brought up, one of the things I wanted to also mention is you can, completely with the best of intentions, not realizing that you're doing it, also build in bias into your systems, such as the recidivism example that you mentioned. You need to be really intensely aware of what your data set that you're using in order to train your system really reflects, and not only what the data is, how it was collected and how it's going to be used by your system, or you can end up building in these sort of implicit biases into these systems that people assume are trustworthy.

April Galyardt: Right. And so one of the examples-- one of the big tech companies tried to build a résumé machine learning algorithm, that it would go through the résumés and try to find people who were the most successful, but because the company had mostly hired males in the past, the algorithm learned very quickly that anything that was a woman's college or Bryn Mawr or all these things, they weren't like anybody they'd hired before no matter how qualified they were, so the algorithm immediately dismissed them. That company, as soon as they figured out what the algorithm was doing, shut it down, but that's exactly the sort of bias that you can introduce because the algorithm picks up on the data which has how the world is; it's very hard how to tell the algorithm how we want the world to be.

Rotem Guttman: Absolutely. So that's where we get into situations where we have to actually create-- I don't want to say fake data-- but we have to modify some of the data or introduce additional data, look at different data sources. So perhaps we don't look in just the résumés of the people from our company that were successful; we try to broaden the net and we look at, "Okay, what's a representative sample from our industry?" Or, "If our industry has a problem of diversity, what are industries that don't have that problem, and how can we push to them?"

April Galyardt: Right. Pulling in additional data, there's so much extra information there, and that's provided a huge boost in a lot of different cases.

Rotem Guttman: But again, that too is a double-edged sword, is how do we control that data that's-- there is a lot of data out there about you, about me, about individuals, about organizations, and that individual in that organization doesn't always control or own that data, and so it can become a problem. How do we express to the users or the people that are going to be affected by these AI systems, "Hey, here's the data that we used to make this determination," when that's not always clear from the system?

April Galyardt: Absolutely. Even knowing what data was pulled in and just-- an example of like who has your data and what they have-- so if you get emails from certain companies you like to do business with and you click on something, you start seeing ads everywhere. That isn't necessarily because Facebook knows that you clicked on that ad, but the company sold your email to Facebook, and so now Facebook knows you did business with that company and they can match it up with other-- and my point here is that we don't even necessarily know how these bridges are being made, and I think that is something that we know regulation is coming; we don't necessarily know what shape it's going to take yet; but I think that, "Who has my data and where is it even going?"-- that's going to be something that it has to address.

Rotem Guttman: Absolutely, and that's a place where really the research can inform the policymakers to make sure that we're making good, relevant policies that are going to help give the effects that we want to see from these technologies as they mature. And I've seen a lot of calls recently from-- not just from politicians but also from industry-- where people are saying, "Look, we need to regulate this now, because unregulated, there's going to be a lot of-- we're really not sure what the world we're walking into is going to look like."

April Galyardt: Well, I think that's exactly true-- we don't know. The way that information was spread during the last election, I think that surprised many of us, and so the point that we don't necessarily know what we're walking into I think is absolutely true.

Rotem Guttman: The last thing I just kind of want to ask is: What's your most hopeful outlook for artificial intelligence? What is your hope of what we can do with it? Optimistically.

April Galyardt: Optimistically. You know, it's funny, I was talking with somebody the other day, and I was like, "How far are we away from an R2D2 unit, like a personal R2D2 unit, for everybody?" And we got to thinking about it, and the answer's really not that far, because we've got the maps, we've got the echoes and all of these little pieces, all of the ways in which you have a personal assistant who can help you. That's all coming together, and so that's maybe the soonest nice outcome.

Rotem Guttman: I'm just hoping that I end up with an R2D2 and not Marvin, so.

April Galyardt: Yes. Yes.

Rotem Guttman: And so I think-- if you'd like some more information on artificial intelligence, please contact us at info@sei.cmu.edu, or click on the link in the description, and we'll be in the chat after this video and during. If you'd like to ask any questions, please feel free.

Related Resources

AI at the SEI
Ginger

Secure Your Code with AI and NLP
CMU AI
Deep Learning and Satellite Imagery: DIUx Xview Challenge
Deep Learning in Cybersecurity