Kira Berman: Welcome, everybody. Can I get your attention in the center of the room, please? My name is Kira Berman. I'm the assistant director for education at the Museum of Natural History. I have been organizing science cafes just like this for more than 15 years here at Connor O'Neill's. Please thank Connor for all the wonderful work that they've done.

[Applause]

Kira Berman: I have a couple of brief announcements about the Museum of Natural History before we begin tonight's program, but I'm very excited about our topic of artificial intelligence and machine learning. One of the announcements is that the Museum of Natural History will soon be beginning its summer hours. Right now, we're open 10:00 to 4:00 on Tuesday through Sunday. We will be open starting in June 10:00 to 5:00 daily, every day, so more time for you to enjoy the museum. There's tons of things to enjoy, including our current special traveling exhibit about dinosaurs, our favorite topic. Please come and check that out. It's from the American Museum of Natural History.

In addition, I have a secret. This is the first time I'm mentioning this publicly, and I don't have a date when this will begin, but we are working very, very hard to have laser rock and roll shows in our planetarium. You heard it here first. That's right. Now you know and coming soon. Coming soon. This evening's Science Cafe is sponsored as we have one of our Science Cafes generously sponsored every year by Sigma Xi, the Scientific Research Honor Society. Gus is here to tell us a little bit about that.

[Applause]

Gus: Thank you, Kira. Thank you very much, Kira. I'll be very brief. Sigma Xi, for those of you who don't know it, is the world's largest multidisciplinary honor society for scientists and engineers. It was formed in 1886 when Phi Beta Kappa refused to elect scientists and engineers into their honor society. The University of Michigan chapter was formed in 1903. At this point, the chapter has more than 1,000 members at all levels. The mission of the society is to enhance research, foster integrity in science and engineering, and promote the public understanding of science.

It's the latter mission, the promoting of public understanding of science, that motivates our chapter to sponsor events such as the tonight's Science Cafe. Other activities of our chapter include judging and awarding prizes at the Southeast Michigan Science Fair, and giving cash award to Science and Math Teacher of the Year, whose teaching inspires, stimulates, and challenges students. Thanks for giving me a chance to explain Sigma Xi and our mission.

[Applause]

Gus: Thank you.

Kira Berman: Gus and all the Sigma Xi members who are here, thank you so much again for your generous and continued sponsorship. Also making tonight's cafe possible is the National Science Foundation. I'm about to introduce our speaker. Before I do, I want to make sure everybody knows our format. What we do is we do a little bit of a presentation at the beginning to give you some input, some information about both the topic generally and the research that our speaker is involved in, and after that, we give you a time to talk about what you've heard at your tables. There are some discussion questions which you may have seen on your tables. The last third or so of our Science Cafe will be a group discussion. I will moderate that discussion. That's our format. We'll have our presentation in just a moment. I'm really happy to introduce Raed Al Kontar.

He's an assistant professor in the Industrial and Operations Engineering Department at the University of Michigan and an affiliate with the Michigan Institute for Data Science. Raed's research focuses on personalized, collaborative, and distributed machine learning, and he will tell you what that means. Raed obtained an undergraduate degree in civil and environmental engineering and mathematics from the American University of Beirut in 2014. He received a master's degree in statistics in 2017 and a PhD in industrial and systems engineering in 2018, both from the University of Wisconsin-Madison. Raed's research is supported by multiple governmental agencies, including a career award from the National Science Foundation, which helped to make this program possible. His work has also won 12 Best Paper Awards in the past four years. Please welcome Raed Al Kontar.

[Applause]

Raed Al Kontar: Okay. Thank you, everyone, for joining us today. I'm Raed. I'll be talking to you today about basically what is AI and what is machine learning? This is the topic today. Just as a disclaimer, I'm a statistician by training. Some of the topics will be slightly technical. I try to make the talk as much applied as possible, but because machine learning in itself, as you will see in a bit, is a mathematical tool. Let's get started. For the outline today, I'll be talking about what is machine learning? Then I'll give you some examples of machine learning. Then I'll discuss what is artificial intelligence and give you some examples, and I will end by discussing the risks. AI and machine learning come with a lot of risks.

Many of us think that they can do a lot of things that, in reality, they cannot still, and we are very far away from achieving real artificial intelligence. Let me start by asking this question. Any of you have used ChatGPT recently? Awesome. Any of you does not know what ChatGPT means? Awesome. I'll highlight a little bit as I proceed what ChatGPT is as well. Let's start with this fundamental question. What is ML? This is through a search through the internet. I did the search.

This is some definitions that you can find on Wikipedia and multiple websites, and you can find that those definitions are very similar. ML, machine learning, is defined as a form of artificial intelligence that enables a machine, a computer, a tool to learn from data to make predictions and decisions rather than through explicit programming. What is the underlying idea here? It's we are giving data to a machine, and the machine is taking this data, extracting knowledge to do better predictions and better decisions.

That's it and the key point here across all the definitions that you have, for example, let's look at the last definition. Machine learning is the study of computer algorithms that improves automatically through experience and by the use of data. All the definitions of machine learning are fundamentally centered around data. Machine learning, in a nutshell, is simply the ability to take data and make informed decisions through the analysis of this data, recognize some patterns from this data, do some predictions of the

future, and do some good decisions. This is what machine learning. I'll be giving you some examples, but what I want you to remember, this notion is machine learning is based on data. This is a key distinction between machine learning and AI as I will talk to you in a bit.

Data is everywhere. ChatGPT is based on data. It took this large language model that can write texts, that can write stories. It is based on all the data on the internet, all the texts, all the novels, all the research papers on the internet. It was trained. It was given this data and based on this data, it is able to write, adjust, improve grammar, correct spelling mistakes, write long texts. This is what ChatGPT is based upon. Autonomous vehicles, self-driving vehicles, they are based on data.

How do they drive in real life? A self-driving vehicle is moving, and the sensors are collecting images from its environment and based on these images or the data, it's making decisions. Should I turn right? Should I turn left? Should I stop? What is the optimal decision that I should do? This is why data is everywhere these days. Really, data holds a lot of power in the modern world, and it has given us this ability to analyze and improve data.

This is an example. This is an example of an autonomous vehicle driving. As you can see, as it's collecting data through this imaging system, and based on the imaging system, it's deciding, “Okay, I should turn right. I should continue straight. What should I do?” The same thing as the robot. The robot is collecting data from its environment, the wind, the temperature around it, okay? Given its history, it was trained to stand on one leg. Those are some examples of the success of machine learning, but is it always successful? The answer is a definite no.

For example, as you can see in this situation, the robot can easily fail. Self-driving vehicles at this stage can easily lead to fatal errors, to deaths in many situations. We are still developing a lot of the tools in machine learning. This is the promise of machine learning. It is somehow autonomy. It is the ability of the machine to take data and make decisions on its own. This is the car. It's moving left or right on its own, okay?

In general, when someone says, “Okay, what is machine learning? What are the branches of machine learning?” It's using data. That's a good definition. What are the different avenues within machine learning? In general, those avenues can be separated into three avenues. The first avenue is descriptive. Descriptive machine learning. What is descriptive machine learning? To give you an example, if you have a new medication, you cannot mathematically prove that this medication will help, will decrease cholesterol levels, for example.

What you can do is you give this medication to patients. You collect their cholesterol levels, and then do some analysis, some statistical analysis, to find whether indeed there is an impact or not. You analyze the data, and your goal was a descriptive purpose to say whether this medication can or there is enough evidence that it can decrease the cholesterol level or not. This is a descriptive problem.
Then there is the predictive machine learning. The predictive machine learning is a setting where you're trying simply to predict the future. For example, companies like Amazon, they are always trying to predict demand such that your Amazon packages can be delivered within a day. This ability to predict demand accurately is very important. This is an example of predictive machine learning. Let me move a slide of some work that I have been doing with General Motors.

General Motors, any of you have heard about the OnStar system with General Motors? The OnStar system is simply a system that collects data from your cars on the road in real time, if you're registered for the service. The hope of GM is to keep the drivers informed about the health of their vehicles, okay? This is an example here on the screen where data in real time from the car batteries, from your car battery, is being collected. Then we try to predict what is the remaining life of your battery through the data, through machine learning. By predicting the remaining life of the battery, this brings us to the next stage.

We have this belief right now that your battery will fail in the next three months, but we're not sure that it's exactly three months. We believe it is maybe between four to six months. The hope then is to say, “Okay, when should we alert the driver? What should we do? What is the best action to take?” This leads to the last part, which

is the decision making, or the prescriptive machine learning. The prescriptive machine learning is we take the predictions, but the hope is to do decisions. One very famous example of decision making in machine learning is movie recommendations.

When you open Netflix, you will find that Netflix will tell you, "Okay, I recommend those specific movies for you." What did Netflix do? It learned your patterns, what you like, what you dislike. Based on their patterns and how you browse the internet, it made a decision what movies should be recommended for every single person? This is an example of how data is being used to describe, to predict, and prescribe. Those are the avenues of what machine learning can do.

I wanted to emphasize how do we learn those models? Just to give you a general idea about how to learn, because behind any machine learning algorithm is a model, is a mathematical model, is an algorithm. This mathematical model, given the data, we need to give the machine the data and train it to identify the data. For example, you guys have your phones that some smartphones will use facial recognition. They are trained, when they see your image, to open the phone. They need to be trained. How do we train that?

This is a simple example where, for example, in this situation, we give the machine two sets of images. One set of images is for Leo, and we tell the machine, "Those images are images of Leo." Then we give the machine other images, and we tell the machine, "Those are not images of Leo," and then the machine—then, we train the mathematical algorithm to somehow separate, to be able to distinguish, okay, this is an image of Leo; this is not an image of Leo. Automatically, when the machine, when I train the machine to be able to distinguish those two images, anytime I get a new image of Leo or of not Leo, the machine can automatically tell me, "This is an image of Leo," or "this is not an image of Leo."

This is another example of, for example, how autonomous vehicles work. They take images of the system as you proceed, and every time they classify this image. Is this a pedestrian? If it is and they classify it as a pedestrian, they stop right away. If it's not a pedestrian, and it's a road, they make the decision accordingly. There is always this notion that I give the machine data, but then I train it to do the task that I want. This is how I train it. I give it data

of what I want it to classify, and what I don't want it to classify, and I make sure that mathematically, we are able to learn the separation. There is fundamental challenges these days with machine learning, is these days we have two curses.

There is this curse of dimensionality. When you have a lot of data, when you have a lot of dimensions, systems like autonomous self-driving vehicles are very complex. Sometimes, for example, to give you this example with GM, GM gets billions of data points every second. Training machine learning models with all these data points, at some point can become impossible, can become very tedious. This is a very fundamental challenge, is training the models and being able to learn from all this data is not an easy task. This is the curse of dimensionality.

Interestingly enough, there is also in many situations, a curse of rarity. For example, in autonomous vehicles, these accidents, the bad situations, are very rare. If a machine wants to learn what constitutes an accident, what should I do, it should have a lot of information from bad events. If a machine wants to learn, “How does Leo look like?” we should need to give it a lot of images of Leo. In many situations, accidents are rare events. Although you have this big data, but in many situations for the events that matter, you do not have a lot of data.
Those are some of the challenges that machine learning indeed faces. This is machine learning. Just remember the take-home message is machine learning is a function of data. What is AI? This is our next question. AI is a subset of machine learning, okay? AI is a little bit more general than machine learning. Those are some of the definitions of AI, and put it this way, AI has a very broad definition. There is no specific definition, “What is AI?” It's just, in its definition is, it's a tool that enables machines to simulate human behavior, okay? How is machine learning a part of it?

When data is involved, when we use data to train a machine to replicate or simulate a human, this subset becomes machine learning. AI is a more generic definition. It has a broader definition in the sense it's—the goal of AI is to make a machine capable of solving complex problems like humans. Then again, if we give data to the AI and tell the AI—for example, we give data to the phone and teach the phone how to predict the next word, to predict it like a human being, this becomes machine learning. This is the

key distinction. AI is this ability to learn, reason, and self-correct with time.

Machine learning are those abilities, but only when trained using data. When data is given to a model, and we train a machine to learn and to self-correct, then this becomes machine learning. Again, machine learning doesn't have the ability to reason. This is another challenge with AI. To give you an example, what are some examples of AI that are beyond machine learning, that are not trained on data? For example, chess.

When you play chess online, when you play against a computer, for example, the computer has not been trained on data. Instead, it has been trained on rules. It is basically the rules are if-then rules. If you move like this, then the computer should do this. If you move in this way, then the computer should take this action. Those are not trained on data, but this rule-based algorithm, or logic-based algorithm, is an artificial intelligence algorithm aimed at replicate what a human being can do.

This is one example of AI that does not fall within the realm of machine learning. Another example is some of your smart appliances at home. For example, some of the battery, some of the bulbs, when you go outside the room, they turn off. When you come inside the room, they turn on. Those situations were not trained on data, but they have this rule, this logic, that if the sensor catches movement, you stop. If it doesn't, you turn on. If there is no movement, you turn off. This is one example of AI that is not machine learning, but really, what the message here is, is machine learning is only a subset of AI.

AI is a general category of teaching a machine to learn, reason, and take actions, take smart actions, but when we give the AI data, and through the data, we teach them how to do such tasks, this becomes machine learning. This is really the fundamental distinction, but here, I want to switch a little bit gears and talk about what are some of the risks that AI comes with and machine learning as well? The first risk is that AI in general, as of now, is a black box, okay? What does a black box mean?

We really do not understand how—what are the underlying principles exactly that would enable this machine, when given a

specific image of Leo, to tell us it is indeed Leo? We still don't understand it exactly. The problem, beyond that, we do not understand. For example, let me give you this example here. Let's assume a self-driving vehicle was trained, was given a lot of images of roads that do not have snow. We trained this autonomous vehicle to drive on places that do not have snow. Suddenly, we take this car to a place that has snow. It's seeing images it has not seen before. What happens then? We do not know because it was not trained on such data.

This is the fundamental issue, and it's a black box. We do not know how things interact with each other. This becomes a very fundamental problem is, what if the machine learning sees something outside its domain? What happens? This is the difference between a human being. We can reason; we can learn outside what we have been trained on, but a machine learning algorithm cannot.

There is another issue of fairness and bias. Let's assume you're doing image recognition and based on—Apple is doing some image recognition. This has been an issue, by the way. These image recognition algorithms, what they do is they take images of faces, and they try to identify whether this person is the right person opening the phone or not, but in some situations, based on the color of the face, based on the color, because some groups are majority groups and some groups are minority groups, if you give it a lot of data from one group and very little data from another group, the machine learning algorithm will do well on the majority group, but will do extremely bad on the minority group.

This is why, in many situations, those facial recognition algorithms were not working for certain ethnicities, and this was a big issue in fairness and bias. Then there is safety. This is an example of Waymo, which is a self-driving vehicle. There was a person dead recently with this accident. Basically, it seems in this situation, the car saw a situation that it has not seen before and made the wrong decision, and basically, it hit the pedestrian. This is another example of safety concerns. There is a fundamental problem that I always tell people: the math. We tend to overestimate what machine learning can do.

The math is still not there yet. We are still very far away from really solving some of the fundamental problems in AI and machine learning. We are far away. Although there is a lot of progress, but there are some problems that are very fundamental, very fundamental in solving that we still are far away. Some of these problems, unfortunately, we have proofs that they're not solvable, and that's a problem.

We are pushing a field that we cannot, it seems, in some angles, given our knowledge right now, we can prove that, given our knowledge, we cannot solve it. What happens then? Then there is some fundamental risks and decisions, going back to General Motors. One example is, if we can predict, if we can predict when failure will happen for your vehicle. For example, I can predict that failure of the battery will happen in between five to seven months. What action should I take?

This is a tough question to answer because I can be conservative, and I can ask the driver, “After four months, come replace your battery.” That means replacing the battery very often, and that will be costly, but then I can wait a little bit of time, but then I run the risk of an accident. This is also costly. How to do decisions? What is the optimal decision to do in this situation is often also unclear. I will give you an example along this line on a project that I have been working with. We have been working with—we have been working with the National Institute of Health on a project where, in pharmacies, in many cases there are medical dispensing errors.

Wrong medication are placed in the pills. This happens a lot, and it's one of the leading causes of death from a pharmaceutical perspective. What we are trying to do is, we're trying to put a machine that after the pharmacy, a pharmacist fills in the pills, the machine will take images of the bottle and try to learn or identify whether there is a wrong medication or not. This is the hope. This is what we are trying to do.

Basically, we have this machine and we're training it to detect wrong medication dispensing, and this is our machine learning algorithm. Actually we developed something, a platform to show the pharmacist. We tell the pharmacist, in this situation, we tell the pharmacist that the machine believes that you did the right job. This is how confident the machine is, the AI is, that you did the

right job. It turns out that the machine can be super confident but super wrong. It's a rare event, but it happens. It happens, and this becomes a problem.

This is very risky because although the machine is making mistakes 0.1 percent of the time, but if it puts Tylenol instead of Advil, that's fine, but if it switches two very dangerous medications, that's a big problem. When to use it? How to use it? Is it even usable is a very big problem for us that we don't have a solution till now? Here I just wanted to emphasize that AI at this stage comes with its own risks.

Us as statisticians and people working in this area are trying to solve some of them, but I can assure you some of them are still far-reaching, but that said, in many situations where risks are low, we have had a lot of progress, a lot of amazing progress as of now. Indeed, AI machine learning is the topic of the decade. By this, I'll end my talk. Thank you, everyone.

[Applause]

Kira Berman: Alright. Thanks, everyone. At this point, we'll take a break for you to discuss what you have heard. Raed will walk around and you'll get a chance to meet him personally, hopefully, as he circulates the room. We'll come back together in about 25 minutes or so for a group discussion, and I'll make sure the wait staff gets in here, too. Make sure you remember the wait staff. Thanks.

Male Voice: Raed.

Kira Berman: Yeah. Do you have a question, Peter? Okay.

Audience member 1: I have a question. Could you please explain?

Kira Berman: Speak into the mic.

Audience member 1: Could you please explain whether it was machine learning or artificial intelligence in four fields of human—four fields of endeavors? One, human robotics, humanoid robotics, and driverless car.

Kira Berman: What's that?

Audience member 1: Driverless car. Search for drugs in medicine, medical discovery and evolve, in general, the process of acquiring data, analyzing it, and learning from the data and make better decisions from it.

[Extraneous conversation 00:31:28 - 00:31:32]

Audience member 1: Could you please explain the reason why?

Raed Al Kontar: I think the question was to explain a little bit the use of AI in robotics, in self-driving vehicles, and in drug discovery. I spoke a little bit about self-driving vehicles in the presentation, so maybe one thing that I will focus on is this notion of drug discovery. Actually, drug discovery and robotics, so drug discovery is actually a very active topic in machine learning, and these days, we have a very interesting notion. I was just having a conversation that I think was very interesting. In the past, so I'm a twin. In the past, they used to do a lot of experiments on twins.

Why is that? They wanted to somehow fix all the other attributes, change one thing, and see what happens, and this is, in statistics, we always have this notion that correlation does not imply causation. For example, one can say that my vocabulary and the health of my teeth are correlated. The reason, it's not that if my teeth become worse, then I know a lot of words. There is no causality between them. Basically, there is another factor. As I grow older, my teeth can get worse, but at the same time, I get more vocabulary. I learn more words.

This is an example of why causation does not cause—correlation does not imply causation, and this is why people do a lot of tests on twins, because they want everything to be fixed between the gene makeup, and one thing is changed. One advantage in drug discovery is if you have access to all the data, from all this historical data from the different healthcare providers, if they are willing to share the data, then what we can do is somehow we can identify individuals with similar attributes, with similar life habits. By looking at those subsets, we can look at causality. We can understand a little bit.

We can move from our history of learning correlations in medicine to perhaps identifying interesting causations. This is my take on this issue that I think is very promising. From robots, indeed

machine learning, if robots are moving, the movement of robots, how they move, highly depends on machine learning. Robots, when they are trying to navigate this room, how do they navigate this room? They have some sensors, and they have some cameras. It takes data. It collects data from the environment, and in real time, it makes decisions.

First of all, it does predictions, because it predicts what is the path to reach the door, for example, and then it takes decisions, as it moves, if someone passes by it, it sees someone passing by, then it makes decisions. "Okay, I need to stop," or "I need to continue moving." Machine learning is this ability to use this data in real time and make smart decisions. This is my take on the questions you raised. Maybe if someone else has something to add, this is a collaborative discussion, please feel free to do so. Okay.

Male voice: Well, okay.

Kira Berman: *[Unintelligible 00:35:02]*.

Audience member 2: Is this on? Okay. As regards to drug development, let me propose a different approach, or one that I was somewhat familiar with. Compounds that result in some medical benefit have a certain composition. If you analyze that—and the way that they develop new drugs is they look at a compound, and they try it. They give it to something or somebody with that condition and see what happens. They have to try a lot of different compounds. If artificial intelligence would analyze all the existing beneficial drugs, and then look for other compounds that had those same constituents, you'd limit the number of compounds you'd have to try in order to see what the effect was.

Raed Al Kontar: Excellent point but to tell you something, this is being done. This is being heavily done. One, to give you an example, just to take it from, because this notion of deciding on the compound is like you're trying to bake a cake, and you're trying to find what is the optimal amount of sugar to get the best cake. The hope is really to get to the best amount of sugar with the fewest experiments possible. That's what you're hinting to.

I want to find the compound that gives me the best outcome with the fewest experiments possible. This is actually a very big use

case of machine learning. Machine learning is very useful in those situations where, because, think about it in this way. If you were trying to find the optimal sugar level before machine learning, what people used to do, they say, “Let me do an experiment with zero grams of sugar, two grams, four grams, six grams, eight grams, 10 grams,” and then choose the best. You did five experiments. Then if you have sugar and salt, then you need to do five square, 25 experiments.

If you have sugar, salt, and flour, you can see the number of experiments becomes extremely huge, and the complexity of the drugs requires a lot of experiments. Machine learning is able to reduce the number of experiments by a lot. You can actually do it with much less experiments. People are doing this a lot in pharmaceuticals, in computer science, and actually, very recently. This is an example why I give this, making a cake, because people were able to try—were using a machine learning model to actually find the best cake. They did that with very few trials. This is an actual experiment people have done.

Kira Berman: That's awesome, but is more cake or less cake better?

Male voice: More cake.

Kira Berman: I had a question from this gentleman over here.

Audience Member 3: Okay, thank you. I'm not sure about the answer to this, but you mentioned chess or go playing programs or machines, as being rule-based. That is, of course, there are rules to the game. I think, my sense is that, let's say, the two I can think of is Deep Blue and AlphaGo programs. I've gotta think that they were trained on past games, and maybe also their own play of games, in order to learn, in order to get better. They weren't programmed to already be better. They had to learn that through machine learning, but, of course, the rules of chess were just part of that program, the initial program. I mean, does that make sense? Is it true that those did use a form of machine learning to train the machines?

Raed Al Kontar: Yeah, okay. The question is about the use of machine learning in games like AlphaGo and chess. Indeed, for example, AlphaGo was trained using machine learning. Actually, right now, no player can beat the AI in AlphaGo. The same thing in chess. No player can

beat an AI in chess. No player. Actually, Magnus Carlsen, I think he's the world chess winner. He has not been able to beat an AI for the past 10 years. Ten years ago, it was impossible to beat. In those situations, just to highlight that, I mentioned the use of rule-based algorithms. In many games that you play, actually, chess can be done in two ways. One is just rule-based. You teach it. If something happens, you change, but more recently, people have been using machine learning.

One beauty of those specific situations is the following. This is what makes it plausible. In the time that you play a single game of chess, the computer can play more than a billion. You can imagine why it can learn much faster than a human because it plays a lot of games, and it learns those policies, what they call policies. This is a field of machine learning called reinforcement learning. It learns policies such that it learns, "If this player does this, what are my optimal, not only next move, my next 30 moves?" In general, a chess player, the best chess player usually can think about maybe 12 to 13 moves ahead. I can think maybe three to four.

Those artificial intelligence models can think usually 30 to 50 moves ahead, at least. This is why it's impossible to beat them, but you are right. AlphaGo is definitely trained on a machine learning model. It's not rule-based. This is why it is part of a machine learning model, but if you're on a plane, you're playing the chess on a plane, usually those are not ML. Those are usually rule-based.

Kira Berman: That's a good point, yeah. I think this gentleman and then Margaret.

Audience Member 4: Okay, back in early machine learning, like the welding machines that building cars, they just did a program, and they had no self-awareness and occasionally they'd hit people and there'd be injuries. Now we have the Waymo's driving around Ann Arbor. I went down for the Hash Bash to drop off a friend and the students walking across the street wouldn't let us drive when there was a green light. Traffic was backed up down there forever because of student behavior, but the risk for students with a human is we might hit them. With AI, if they can't hit a—you could put out cutouts of people and you could stop a car in a bad neighborhood, and you could strip the Waymo of all of its 40, $70,000 worth of

machinery. Will you climb in a self-driving car and drive through a bad neighborhood?

Raed Al Kontar: I'm not sure how to answer the question, but I mean, yeah, I think to your point, many situations, AI can do better than the human. Many situations. Actually, in that example of the image dispensing, the error rate of the AI is less than 0.001 percent. The question, will it be implemented? That's a hard question because who takes the blame if something bad happens? Is it the AI? Is it the people who developed the AI? Who is it? This is a very fundamental question, and put it this way, Tesla and those self-driving vehicles, they have this—they claim it's a level five automation, but they always require you to have your hands on the wheel.
Why is that? They know they cannot guarantee the AI. They know that they cannot tell you, "Just base all your decisions on the AI." They know that if something happens with your hands on the wheel, you are the ones to blame. This is why this is an important thing because, again, the math is not there. We are not there yet, and this is why it is still very important for the companies to have the human taking the final action. Maybe things will change in the future, but right now, I always mention that we are not there to guarantee that we can always be accurate, but at some points, we need to ask this question on how much risk are we willing to take and what is too risky for us? I hope that answered the question.

Audience Member 5: I want to give an example of humans that beat robots or actually, that beat AI or anything else. Most of us here, I suspect, are grandparents. Have you observed your child learning language? In one instance, they can learn words. You would have to take, at least on current models of statistical regularity, thousands of instances for one word to be learned, and that kid has it down. It would suggest you have the wrong models or the wrong algorithms for your AI models.

Raed Al Kontar: I think this is a fundamental advantage of human reasoning compared to a machine learning model. By definition, those machine learning models can only learn with a lot of data. This is a fundamental problem, and they can only learn within the bounds of the data that they acquired. If I teach you to separate, as I gave this example, if I teach you to separate a dog from a cat, and then you give the machine learning model a picture of an alligator, it doesn't know what to do.

A human being knows what to do in these situations, and this ability to go beyond our current boundaries of what we were trained on is really a key distinction, a key ability that people in the machine learning community want to get to this level of global knowledge, but we are very far away from that. This is a fundamental disadvantage, fundamental limitation, as of now of machine learning is first of all, the need to only learn. The machine learning model can only learn within what it was taught to learn. That's it. It cannot go beyond. This is one thing, and the other thing, I gave this example about self-driving vehicles, you need a lot of data.

If you want to avoid accidents, you actually need a lot of data from accidents to be able to detect that. This is why a machine learning model is only dependent on having a lot of data. Let me give you an example. ChatGPT, the model for ChatGPT, models for natural language processing have been available for more than 30 years, but only today they have become plausible because only today was someone able to look at all the internet, get the data from all the internet, and train the model on the entire data. This ability to train with all this data is what made it possible. Not only the math, the math advanced recently, but we had preliminary models that could have done a very similar job.

Audience Member 6: Okay, if you have, I'm talking about now, the future. If you have people who have suffered brain damage, accidents, tumors, you name it, and they lose their ability to think clearly, do you foresee some kind of interaction or something put into that person that would replace that function?

Raed Al Kontar: That's a tough question for me. Put this way, I'm a mathematician. I spend my day doing proofs, so I am not sure. I hope so. I hope there is. Hopefully machine learning can replace cognition, for example.

Audience Member 6: *[Unintelligible 00:47:24].*

Raed Al Kontar: Yeah, but at this stage, I have to say, at this stage we are very far away from this situation because, again, going back to the previous question, to be able to reach that level, the model should be able to

learn beyond the boundaries that it was trained on, and that's not currently possible. Yeah.

Audience Member 6: Will it be possible?

Kira Berman: Will it be possible is the question.

Raed Al Kontar: I'm not sure. It's not done yet, but based on the past decade of changes, there is a probability. How big the probability is, I think your guess is as good as anyone here and my guess, so I'm not sure, but I think the rate of change in machine learning has been extraordinary in the past few years, so I think things will move forward, but still I have to say that there are some fundamental knowledges in the math. If those are broken, I imagine the slope of increase can become super large.

Female voice: *[Unintelligible 00:48:37]*.

Audience Member 7: You mentioned a couple of times now the example of if you wanna deal with machine learning on auto accidents, you need a lot of auto accident data to train on. Does that not lead to a perverse incentive for those developing self-driving vehicles to have lots of accidents to gather data on?

Raed Al Kontar: That's a very good point. Yeah, this is a problem because in that situation, people may be incentivized, the companies may be incentivized to have accidents. That's a very fundamental problem, but what are people trying to do? People are trying to use generative AI in the sense they're trying to say, okay, given those 50,000 accident incidents that we have, can I try to generate things that are not the same as those 50,000 but look like them? Can I fabricate instances that are artificial, but can somehow be representative of situations that may happen in the future? This is the current policy a lot of machine learning folks are taking. Hopefully, they don't take the policy you mentioned. This is what people are doing, doing this generative AI, which is basically much like what ChatGPT does. It generates language. They're trying to generate scenarios.

Audience Member 8: Hi, I wanted to follow up on your earlier point on the learning is only as good as the data put into the program. Specifically, I was interested in if there's any knowledge base out there of how high

quality the data is on some of the conclusions we're seeing published. I'm gonna give you one quick example in my life. I use Google Maps to navigate if I go anywhere far away. One thing I have noticed, if I put in a destination in Ann Arbor, I don't know who's figuring out how you get there, but they are going the long way around.

It's like they don't know how to drive in Ann Arbor. I will ignore what it says and I'll beat its time by five minutes going, Waze, Google Maps, whatever one you wanna use. My theory was, oh, the only people that use this are out-of-towners, and they're all taking the main roads. They don't know how to drive around traffic, and so it has no data.

Raed Al Kontar: That's a very good point. There is this famous saying by a famous statistician, George Box, “Rubbish in, rubbish out.” You give the model bad data, it will give you bad decisions, okay? The hope in general is not only to have good data, but it is also, if there is bad data, to detect the data is false; it is misleading. This is an issue. For example, to give you an example, Google, if you have a Google phone, most recently Google is trying somehow to learn models on your phones. If you have the phone in the first few weeks, it may be slower than usual because Google is using your phones to learn some of their ML models.

They found that some—so basically every phone does an update and this update is sent to Google and Google learns those big models by using the computational power of the phones. They found that a lot of the updates from some people can be misleading, and it turns out that there are some people intentionally sending wrong data, bad data, that really ruined their models. This was a big issue for them. How to address the situation? They tried to develop some ways to detect anomalies, detect what they call adversaries, but still, this is a big issue. Again, any model is as good as the data it was provided. Yeah.

Audience Member 8: *[Distorted audio 00:52:48]*.

Kira Berman: Oh, she says, what about incomplete data?

Audience Member 8: The problem is that I know how to drive Ann Arbor. The problem is if I know how to drive Ann Arbor, I'm not consulting Google

Maps. Everyone knows how to drive it, doesn't need it, so they have a very biased set.

Raed Al Kontar: This is another problem that I mentioned is the bias in the data. In your situation, the people giving the data are the people that are not experts. This is a fundamental issue, and then in general, for Google Maps, so they don't only depend on data, they actually—some companies send cars just to learn the roads, just to improve the quality of the data. If you just depend on some not very accurate imaging techniques or feedback from people, sometimes you will be prone to big errors.

Audience Member 9: This sort of ties into that maybe. Can you say something about prioritization of—it's not always, "Well, this is the right way and this is the wrong way." Say, a self-driving car, for example. Okay, I'm in this situation now. I have a choice of hitting this pedestrian, probably injuring or killing them, or hitting a tree and injuring or killing me. How do they deal with that?

Raed Al Kontar: Yeah, this is a tough question, again, because this is a decision that policy makers should do. This is a decision that the ML—this is an input to the ML. This is the issue, that you can tell them, of course, the ML doesn't mean it will be super accurate and abide all the time, but then the policy makers should decide on what is the risk? What is tolerable risk? Hitting a pedestrian, hitting the car, what should be done in these situations? This is why, and then again, to go back, many companies are not willing to take the risk yet.

This is why, going back to the autonomous vehicles, all the cars, if you have Teslas, if you have any self-driving car, autonomous car right now, you are always required to have your hands on the wheel, on the driving, because they cannot take the risk as of now.

Kira Berman: There we go. I'm reminded of the laws of robotics from Isaac Asimov. Yes, yes, I'm that old.

Audience Member 10:

Continuing on with the self-driving cars, over the course of the last 50 plus years that I've had driver's license, I've had a lot of times where I've driven 400 or 500 or more miles in a day. I would love to have a self-driving car, but if I put my hands on the steering wheel and keep it there for the next eight, nine, 10 hours, I know at

one point, at some point, there's gonna become a disconnect between my brain and my hands, because I'm a human being. Just because the hands are on the steering wheel doesn't necessarily help.

There's gonna be this degradation in the ability of a human being, and I suspect there's probably gonna be a quicker degradation of my ability when I have my hands on the steering wheel, but I'm doing absolutely nothing beyond that, than it would be if I was actively driving. How do they handle situations like that?

Raed Al Kontar: Again, let me just answer this from one perspective that's slightly different. They may want your hands on the driving wheel, just not to take the blame. This may be the case, and this is the case with high probability. Again, if autonomous vehicles improve safety, improve safety in the sense they can reduce the accidents, but even those accidents that happen because of autonomous vehicles, it's still not clear there is no policy around them. It's still unclear who takes the blame, and before doing their wide-scale implementation, really still needs policy. Again, going back to the example, indeed, it may be the case that if it's fully autonomous, the system is doing better, but this notion of blame is what's preventing the actual implementation.

Kira Berman: They're not actually thinking about whether—what's best for us. They're thinking about how not to take the blame.

Raed Al Kontar: Probably.

Kira Berman: That's the problem with capitalism. Listen, folks, there are little blue evaluation sheets on your tables, and there are little yellow pencils. Yes, that's blue and yellow. That's not a coincidence. Please use the pencils to fill out the evaluations. Please tell me what topics you would like to hear about next year. I have six Science Cafes to plan. Some people suggested last time microplastics, and there have been a lot of other great suggestions, and I always look. We have time for maybe one or two more questions, and this lady here is gonna get us started.

Audience Member 11:

I'm curious. You've been talking about who you work for and the fact that it's automobile companies and different things like that. What's the military doing?

[Background noise 00:58:51]

Audience Member 11: Sorry.

Raed Al Kontar: I'm not sure. The military is using a lot of the machine learning appliqué. Computer vision is a big part of the research. What exactly they are doing, I don't have any projects with the military specifically, but what we know is that there is a lot of focus on vision and using machine learning to detect adversities much faster than a human can do. I think this is one area that is being used, being investigated. There is also a lot of research going on at the University of Michigan, funded by the DOD, Department of Defense, on the human-machine interaction. If there is an AI and there is a human, how can they interact together in the best way possible, so that their combined output is better than—it's better than their individual output summed together in some sense? The exact type of research, I'm not sure.

The question is what is NASA doing? Again, so NASA, I can tell you a little bit from what they are funding at the University of Michigan. They are funding a lot of research, again, much like the DOD, on how can we do some image recognition? How can we recognize some anomalies? How can we detect anomalies super fast using machine learning and adjust to those anomalies much—very fast as well, to take decisions in real time. For landing, I'm not sure. I think they're more using control theory compared to machine learning at this stage.

This is what I can tell you based on my limited experience with NASA. I have not worked with NASA or the military as of now, but I mean, put it this way, a lot of the machine learning applications that people are using are available. Those are tools, and people can use these tools across different entities. Again, I would say most of the current company uses are more language processing and image recognition. Computer vision, basically.

Kira Berman: Okay, we're gonna take one more question. I'm gonna remind you now, don't forget to pay your bill 'cause otherwise, Connors complains to me. I will take one last question.

Audience Member 12: Thanks. Just a quick comment on that last question. I work for a commercial software company in the AI, ML space. Computer vision is exactly right. That's what the military's looking at. As an example would be analyzing images of the ocean bed to detect shipwrecks and other things. My question was, as far as risk mitigation in generative AI models, what we're seeing now is the application of a lot of other models to analyze a prediction to give you an idea of the correctness, the coherence, the hallucination, all of that. Are you seeing practical use of that in your research?

Raed Al Kontar: Yeah, that's a very good question. Basically, what your question is, people are using models to verify the validity of the generated data, of the generated images. Just let me take this a step back. Those generative models, theoretically speaking, they tell you that. My hope is that the image that I generate or the data that I generate is as close as possible to being real. This is actually how you train the model. When you train the model, you try to minimize the difference between the real image and the generated image. That said, this is not always doable. This is not always plausible because again, the math still has limitations.

This is what those tools are trying to do. They try to classify. They try to build boundaries, decision boundaries, whether this is real, this is fake. Let me tell you something. For earlier on, we had this called generative adversarial networks, GANs, and it was very easy to separate what is real and what is not real. More recently, we have those more advanced models such as diffusion models. In those situations, finding this boundary between what is real and not real is very challenging, and I imagine that as machine learning progresses, as generative AI progresses, it will become much harder to detect whether this text was generated by ChatGPT or by human. This will become much harder and much more challenging. Again, many of these models have theoretical guarantees that if your data is infinite, it becomes impossible to separate. Okay, thank you everyone.

[Applause]

Kira Berman: Thank you, Raed. That was a wonderful, wonderful dive into machine learning. Thank you.

[End of Audio]