CARLSON SCHOOL
OF MANAGEMENT
UNIVERSITY OF MINNESOTA

Human Decisions in the Age of AI

September 19, 2019 at the Carlson School of Management

Algorithms and artificial intelligence make a staggering number of decisions for people and organizations- the products we buy, the news we see, and the people we hire are deeply influenced by these technologies. We were able to recently sit down and talk to leading experts, Kartik Hosanagar and Ravi Bapna, about human decisions in the Age of AI.

Kartik Hosanagar: Machines sort of governing our daily lives. It's pretty significant, and most of us often don't realize it. You get on Netflix, and you decide what to watch. The recommendations are influencing choices we make. A research study by the data scientists at Netflix showed that 80% of our viewing hours are driven by the recommendation algorithms. So that's pretty significant. And if you look at Amazon, over a third of our choices are influenced by recommendation algorithms. At YouTube, 70% of the time we spend on YouTube is attributed to recommendation algorithms. So they make some pretty significant decisions. Of course, these might seem like not-so-significant decisions, but if you look at how people decide who to date, who to marry. When you look at Tinder, Match.com, again, it's the algorithms that are deciding who you meet. So typically, with a lot of our day-to-day decisions, they are being influenced by algorithms. And I was struck by how it changes with age. So I was running a survey with some of my undergraduate students, and one student, in particular, kind of struck me as pretty interesting. I had a whole chapter on this person's typical day.

Everything from what time he wakes up-- so he has a smart alarm that is an algorithm that tracks how he's sleeping and decides what's the optimal time for him to wake up so that he's performing his best during the day. And of course, even things like LinkedIn, who he connects with, where he ended up getting a job. So he had LinkedIn because of a recommendation at LinkedIn to connect with a former classmate. And so they're all around us, and certainly, that's as far as our day-to-day lives are concerned, but if you look at really important decisions at the workplace, life-or-death decisions, those are changing. In courtrooms in the US currently, there are algorithms being used to guide judges in bail and sentencing and parole decisions. Medicine is moving towards precision medicine, and the whole idea there is to use your DNA profile to personalize treatment. So I think that is significant.

Ravi Bapna: Yeah. So I guess, how does this shape who we are as humans, right, in some sense? And your book gets a lot into the tricky aspects of this, right? So there are a lot of unintended consequences. So for instance, I think one of the examples that I picked up early on was this guy who had a big important meeting the next day, and then got a? Facebook notification, which led him to a YouTube video, and then that led him down a series of things that sort of he got sucked into, in some sense of the word, and then he actually probably wasn't

well prepared for this meeting as well as he should have, right? I mean, at some level technology is amazing, but what are some of the costs of having so much of what we do governed by these algorithms, by the tech companies? And is there any kind of work that you've seen or you're familiar with that sort of talks to the unintended consequences of this?

Kartik: Yeah. First, let me state the obvious, which is that these systems create a lot of value, right, because we make these decisions, what to watch on Netflix, what to buy—you are saving so much time that you would have wasted otherwise on these decisions. So a lot of value, but yes, when they're making so many decisions for us, and we tend to be very passive with how we use this technology and algorithms, they do have consequences we don't realize. Some of them are well documented, so for example, I'll start with the obvious ones. Facebook used to use human editors to curate trending news stories. And the human editors, the accusation was they had political biases. They were left-leaning. So they said, "Okay, let's get rid of the human editors. Let's use algorithms." And, of course, they cannot have political bias, but then now, that couldn't detect fake news stories. And so that was an unintended consequence there.

And then I mentioned algorithms used in courtrooms. There was an investigation done by ProPublica that showed that these algorithms had a race bias, and pretty significant race biases. There was a news report late last year by Reuters about Amazon using algorithms for resume screening. And that's a pretty interesting application of machine learning. Amazon hired over 100,000 people last year. They got over a million resumes. So you cannot have human recruiters sort through that many resumes, so makes sense that you want to use some machine learning. But then, their own test found that there was a gender bias in there. And of course, Amazon is a savvy company, so they knew to test for that. They killed it when they realized that was happening, but there are so many companies that are using these--that have no idea what could be going wrong. How do they change things? We are influenced in terms of what music we listen to, what books we read, what we consume, and, after all, who we are is just a function of what we read and all these decisions we make. At the end of the day, everything about us is just a sum of all these decisions, and they're all driven by algorithms. So pretty fundamental. And we're starting to see that they can have many unanticipated consequences.

Ravi: So maybe, for the benefit of the audience, could you give an intuitive sense into how these algorithms can be biased? I mean, people who study this, probably it's obvious, but why would all these algorithms have gender bias, for instance? Just so that we have a common sort of understanding of your work.

Kartik: Yeah, I'm glad you brought that up. I think we should get into it because sometimes people have a misconception about this. In fact, I remember there was an interview with NPR where one of the other guests on the show said, "It's humans coding these algorithms. And these human programmers at tech companies are biased, and they program their biases in." And you and I know that's not the case, that there are people intentionally coding in bias. And

it's different. And maybe I'll back up, and if it's okay, I'll just kind of set up how these algorithms work.

Ravi: Yeah, that'll be great.

Kartik: So if you look at all of these systems, at the end of the day, they are algorithms that are making choices and decisions. And quite simply, for anyone who's not into computer science, an algorithm is just a series of steps a computer follows to get something done. It's as simple as when I make an omelet, I follow a series of steps. I could call it an omelet recipe, but the engineer in me always calls it my omelet algorithm. So that's an algorithm. It's just a series of steps to get something done.

Kartik (cont.): Now if you wanted to build really-- let's say, starting with relatively simple things, like figuring out how much tax we owe this year. If you want to get your taxes done, it's a pretty simple set of rules because those rules are in the rule book. You kind of say, "If this person resides in Minneapolis and their income is over this and they have two dependents, then the rate is this--" and so on. And so that's the algorithm to calculate the tax. That's H&R Block's algorithm, right? Now you want to do something a little more complex. Let's say you want to diagnose disease. You interview a bunch of doctors and say, "How do you diagnose disease?" And the doctors actually give you their rules. So you interview a few hundred doctors, and they give you all these rules. If the person has a fever and the fever has been there for over a week and they have chills and body ache and something else, then I suspect a bacterial cause, and I give them an antibiotic and so on, right? So those are the rules the doctors give us. Now it does sound that, as you go to more and more complex tasks, tasks that we associate with intelligence, these rules don't do so well. You can interview experts, but those rules don't do so well. A simple example of this might be, if I ask you, "Can you recognize your mother's face?" you'd say, "Of course, easy." Then I say, "Give me the rules for recognizing your mother's face." That's tough, right? That's the—

Ravi: Polanyi's paradox.

Kartik: Exactly. It's called Polanyi's paradox, which is we know more, right, than we can tell. There's a lot of tacit knowledge in all of us. So the question is, is there some other way to bring intelligence without coding in the rules? And that other way, which has been around for a while but was never successful but has finally been successful over the last maybe 10-odd years, is machine learning, which is, instead of hard coding the rules, we will observe people make decisions. We will collect the data, and we'll just observe the patterns of behavior. Instead of asking doctors to give us the rules, we will observe a few thousand doctors make diagnoses for tens or hundreds of thousands of patients. We will take that data. The data has details of what symptoms they walk in with. What were their medical markers? What was the final diagnosis?

Now, on the machine learning system is a simple statistical program that's looking at patterns in there and saying that when I observe the person gets an antibiotic, it does solve their fever that's been there for over a week and these other symptoms. So it starts to recognize these patterns, a lot of patterns that we're not able to convey. A lot of patterns we use to recognize faces, but we cannot articulate those patterns, and it picks up those patterns. And that is what is underlying modern machine learning. So most of what is artificial intelligence today, in practice, is machine learning. Artificial intelligence is a broader concept. It's about creating intelligence. If you could do it through the first method I described, which is called expert systems, that is when you interview and get the rules, but -

Ravi: It didn't work

Kartik: It didn't work as well, and so the better approach now is machine learning. So we feed it data, and it figures out patterns. So coming back to your question-- sorry, that was a long explanation.

And there's, I think, a big difference between rule-based systems versus machine learning. I think that's a paradigm shift.

Ravi: That's a paradigm shift. And because of tools, because of big data, because of processing power, we can do this now, right? I mean, we have enough labor data to learn in a way that's practical.

Kartik: That's right. 15 years back, 20 years back, we didn't have that much data. Now we've got a lot of data so you can learn patterns. You cannot learn if you've got not much data. If you've got a lot of data, but you don't have a processing power, that's not-- now you have that much processing power as well. The last thing is fundamental improvements in the algorithm, which is this revolution in deep learning and a technique called back propagation within that, which are algorithmic improvements that allow things to happen. But, yeah, coming to the question of why is it that-- let's say, as to that big Amazon example, why is it that Amazon's algorithm has gender biases? What is being done is you're giving the algorithm data on millions of past applicants. And you're saying these are the people who got jobs. These are the people who got promotions. Now, find people who are likely to hire. Find people who will do well at the workplace. And so if there were passive biases, then the algorithm picks it up. And you would think that the obvious thing to do is you hide gender, right? But it's not as simple as that. You kind of hide the applicant's gender, it looks at the name, and it figures out that this person is female. Then you say, "Let's hide the name as well." Well, even that won't do it. It can look at your interests, and it can look at which college you went to, and the combination of five, six things in the person's vitae is no returning gender.

And it's not that it's trying to predict gender, but it's just correlating all that data. And so it figures it out. So that's where some of these issues arise. And I mentioned, certainly, one kind of unanticipated consequence of that is fairness, which is something people follow. But there's many other issues. There's the fake news issue, which is not a fairness issue. It's a different kind of unanticipated consequence. There is the whole issue of security, and we could talk about that, say, with driverless cars, what's called adversarial machine learning, and so we talk about security. So there's a set of issues, and we can talk over each one.

Ravi: And what are we kind of-- what's this theory we are thinking or practice in terms of correcting for this? Are we going to get to a place where we will have some combination of expert sort of or rule-based systems along with machine learning where-- so how do we start tackling some of these algorithmic bias issues, I guess? Maybe we could stick to that before we move on to some of the other problems you talk about.

Kartik: Yeah, sure. Yeah. And as you hinted, the solutions maintained for a security issue versus-- with the driverless car, you're worried about a failure where it's not seen a pattern in the past data, and it crashes, and so the solution's different. But for fairness, how might we address that? There's a few different ways. So first of all, one way to address it is to explicitly test for fairness.

So for fairness, one of the first things the community had to do was define what is fair. And of course, they found, not surprisingly but also problematic, that there's no one definition of what is fair. And there's actually over 19 different definitions of fairness. And in fact, there's a nice result that shows that these definitions are incompatible. So if you have fairness along one definition, it will not be fair along other definitions. So in each context, you have to define what is the right criteria. And I think in some settings, you do, like the Housing and Urban Development department has a notion of fairness.

And so, given the definition, you can test for fairness, and you can also engineer for fairness. And that's an area of research right now. To my best knowledge, there are a lot of commercial systems that do this very well. IBM, for example, has some open-source solutions to building fairness into your approaches and so on. So it's, I think, reaching a point where it's going from academia to industry. But I think the state of the art there is to formally test for fairness and also engineer for fairness and correct in different ways. Some of it involves going through the data and trying to fix the bias in the data, but mostly, the approach that is getting traction is not to say-- because you don't want to throw away data. But can we actually test for fairness and correct that?

Ravi: So I think this is actually a good time to maybe have you read a small section of your book. And this is one of the sections that I found really fascinating because it also gets into a little bit of your life and how you were sort of thinking about writing during spring break and so

on, stuff that I relate to pretty well. But it also talks about an area, I think, around recommender systems, which maybe we can talk a little bit more about in detail because I think the aspect of recommender system research that you've done, which is different from what a lot of other people do, is to look at, again, actually, some of the unintended or unanticipated consequences of the impact of these recommender systems on the demand solution and other stuff, right? So maybe if you take—I think this is probably, according to my algorithmic timing a little bit, I'm predicting it will take about five minutes, right? But let's see.

Kartik: So let me read this. It's spring break, a rare week on campus without readings, which gives me time to get some writing done. I open Spotify and play my "all-time faves" playlist. It will run in the background over the next few hours while I make progress on this book. I have typed only a few lines when I sense that the music is bothering me. Even though it's been weeks since I logged in to Spotify, I've begun to notice just how familiar all the tracks are. These are mostly songs from the '90s that I've been listening to for more than 20 years. I suppose even all-time faves have a shelf life. I wonder if my YouTube playlist or Pandora stations will be better. A few minutes into each of them, and I realized that every single playlist of mine is dominated by songs from my high school and college years. I'm musically stuck in the '90s.

Because I'm professionally in a good position to be aware of the algorithmic solution to old-fogyism, I decide that writing can wait while I test out, in the name of research, of course, not procrastination, the recommendation engines of three digital platforms: Pandora, Last.fm, and Spotify. I haven't made these choices at random. Each system represents a very different approach, modeled on how we, as humans, might make a recommendation. For example, if you ask me for advice on music, and I knew you liked Yellow by Coldplay, I might try to think of other songs that are acoustically similar to it. Pandora's algorithms are based on this method, known as content-based recommendations.

These systems start with detailed information about a product's characteristics and then search for other products with similar qualities. If I knew you liked Yellow, but I wasn't familiar with Coldplay's music, I might try instead to think of someone I knew who also likes Coldplay and ask her what else she listens to. Last.fm uses the second approach, known as collaborative filtering. Spotify tries to combine these two methods. I know the theory underlying each platform, but how might my experiment play out in practice? I start with Pandora and ask it to recommend music based on my interest in Thunder by Imagine Dragons, one of my rare recent discoveries. I know I make a very sorry appearance of myself, but I have to say I wrote this two and a half years back, so I'm not as clueless. Pandora serves up Ride by the band 21 Pilots. It informs me that the song was recommended because it features, in quotes.

Kartik (Cont.): a reggae feel, acoustic rhythm piano, use of a string ensemble, and major-key tonality. The description sounds convincing enough, and my ears approve. It's not technically a new discovery. I've heard this track before, but this is the first time I registered the name of the

song or the artist. The next recommendation is Weak by AJR, which, Pandora notes, features "similar electronica influences, mild rhythmic syncopation, acoustic rhythm piano, extensive vamping, and major-key tonality." As you can see, Pandora has a deep understanding of music and the vocabulary to articulate it. Its online radio service emerged from the Music Genome Project, in which musicologists listened to individual tracks and assigned more than 450 attributes to each. These range from the obvious, such as the extent of instrumentation in the music, to the esoteric, such as rhythmic syncopation. Does anybody know what syncopation is? Any musicians here?

Audience Member: When the emphasis is on the offbeat.

Kartik: Good. Once you indicate that you like a song on Pandora, the algorithm finds other songs that have similar musical qualities. The first 11 recommendations from Pandora all sound acoustically very similar to Thunder, especially Young Dumb & Broke by Khalid. Trust me, now there's a mashup of Thunder and Young Dumb & Broke laughter. They all met my approval. The 12th track, Sail by AWOLNATION, is okay. Should I give it a thumbs up or a thumbs down or just not rate it? I decide that aggressive choices can better guide the algorithm and go with the thumb down. Pandora now decides to take a dramatic turn and begins playing songs with significant electronica influences. The first two tracks are fine, but I really dislike All Time Low by Jon Bellion. Sorry to this artist, apologies. Another thumbs down, this time without hesitation, and again, the algorithm adapts. I listen to 20 songs in total, 14 of which are completely new to me. The 6 familiar selections aren't ones I would have searched for myself, but I like most of the music. Pandora's algorithms also tell me that I apparently like songs with extensive vamping. I'm not sure what that means exactly, but the next time I don't like a song, I will shout out a request for more vamping. It can't hurt, right?

To try a very different musical genre, I asked Pandora to create a custom station based on my interest in Pashmina, a song from a Bollywood movie composed by the musician Amit Trivedi. I listen to 20 songs it selects, which, it informs me, feature emotional vocals and simple harmonic progressions. I find that the track actually-- I find the tracks to be quite different from one another and suspect the fact that they're all Indian film music dominates other qualifiers for Pandora. Pandora's approach worked for me, but only because someone had taken the time to catalog detailed attributes of all of its offerings. Collecting such data manually is incredibly time consuming and expensive.

Ravi: Can I just-- for those MSBA students who are here, I mean, this is the beauty of feature engineering, right? I mean, these guys have taken this to a level that I think is just sort of unbelievable. And I mean, I don't think there are many other examples of other companies and other genres, like in movies, for instance, right, where they have done something as deep on the content side, right? Is that correct?

Kartik: So the only company that's done that recently is Netflix. So Netflix's recommendation was collaborative filtering, which I'm going to read about in a bit. That was the approach all along. A few years back, they actually did this thing in an amazing exercise. They had just movie enthusiasts watch movies and just write down a few phrases, short phrases to describe the movie. And you get paid for literally watching a movie and saying this is, whatever, Korean anime or something, anything you want to write. So they collected all these attributes, and then they adjusted their recommendation based on this. In fact, a computer science researcher found a bug and was able to extract all of these attributes. So there is this database out there which has - I forget, is it 5,000 or 20,000? - all of these phrases that Netflix uses to describe a movie. Some of them are so detailed and specific, but I guess that's, again, very beautiful feature engineering within that. So coming back, I turn to Last.fm's recommendations. Out of 20 tracks it suggests in response to my entry of Thunder, I'm familiar with only five. Acoustically speaking, some of the songs, like Magic by Coldplay, sound very different from Thunder because the collaborative filtering approach used by Last.fm is based on the "people who bought this also bought that" and "people like you also liked--" kind of recommendations that we often see on Amazon and other websites. While this enhances discovery of a wider range of music, it comes at a cost. While I truly disliked only one of Pandora's recommendations, I disliked about six of Last.fm's recommendations. Moreover, unlike Pandora, Last.fm does not explain why it recommended any of its selections or how they're similar to Thunder. It wouldn't be able to, as it lacks Pandora's depth of musical knowledge. When I try for matches with Pashmina, Last.fm was unable to offer any recommendations. My guess is that it reflects the fact that not many of Last.fm's users have listened to that track, and the site simply lacks the data on which to base "people who listen to Pashmina also listen to--" that type of recommendation.

So yeah, that was it. And I think the one thing I'd add there is I talk about Spotify at the end of the chapter. Spotify combines the two approaches. Spotify started with collaborative filtering and then added the Pandora approach, the feature engineering approach. And now Netflix has also done that, and so many companies are moving in that direction. One interesting thing is that Pandora spent a lot of money-- or time, I should say, paying artists to listen to songs and write all of these. And fortunately for Spotify, there's a lot of unemployed musicians who can sit and listen and write all this stuff. And for Netflix, again, they could ask a lot of people to do this. And what Spotify did is they automated the process. Instead of having musicians listen to songs and say, "This is song has a lot of rhythmic syncopation. That has a lot of vamping," they are able to get machine learning to listen to songs and do that, and so they've automated that process.

Ravi: So a couple of follow-up questions, and then we'll open it up, I think, to the audience. So I think one of the interesting things with the combination of content and collaborative filtering on Spotify was that I think you attributed the fact that, on Spotify, the playlist-- what's it called, all-time favorites?

Discover Weekly. Sorry. Yeah. So this playlist, I think, actually kind of shifts the demand to the long tail, right, in some senses. So will you talk about that aspect of recommender systems? I mean, I think my thinking, when I saw this come out, 15 years ago, 20 years ago or whenever Amazon started, was that this will really push niche products out, right? And I think that was the common belief till you started testing that. So what did you find, and then, I think, why does Spotify sort of-- how is Spotify able to, I guess, have this kind of more just even demand distribution on their [inaudible]? Basically, talk about that, I guess.

Kartik: Yeah, sure. So some of you may remember or may have heard of, at least, if not the book, The Long Tail, you've heard of the concept of long tail. There was a book written by Chris Anderson, who was editor of Wired Magazine. He wrote a book called The Long Tail many moons ago - it was probably 20 years back - saying that the internet will reduce our search costs, help us find products that are close to our preferences independent of their popularity, and therefore, shift the world from a world of blockbusters to a world of the long tail, where niche items can actually thrive. And so, now, when I'd read the book and the related discussions, it made a lot of sense to me. And then in one of my PhD classes, one of my students, when we were talking about recommendation systems and talking about the design of recommender systems, he brought up the possibility that the most popular designs, collaborative filters, they have this, "People who bought this also bought that." So for something to get recommended, it has to be bought by others. So maybe it will not find niche items. It will just find popular items and recommend them.

And so he brought up the possibility. And I said, "That's an interesting theory. Let's test this out." And over the years, I've done a series of studies on this. The first was analytical modeling of this to figure out, "Is that what might happen?" And our modeling suggested that collaborative filters, the most popular recommender designs, will actually create a rich-gets-richer effect because of the bias that I just mentioned. Later I've done empirical studies, and we've validated all this.

Ravi: So you worked with a Canadian retailer, right, and you kind of A/B tested this idea of some people getting getting recommendations using collaborative filtering and others were not.

Kartik: Yeah, so we partnered with a top-five e-commerce retailer globally, and we ran an experiment with their Canada operations. So all of their Canadian users were partitioned into one of two groups, again, randomly partitioned. One group got no recommendations. The other group got collaborative filtering recommendations. And then we observed the diversity of consumption of group one versus group two. And interestingly, we found that the group that got the collaborative filtering recommendations, the diversity of purchases was lower, that the market share of the more successful products was higher with the group that got recommendations. And I think, since then, interestingly, there's been a lot of interest in the computer science community to try and fix these issues. So there's a whole area in computer science now. People call it different things, whether it's adding diversity. Some people talk about

serendipity and so on. But there's new techniques that have come up. But overall, now, people realize that collaborative filters actually have this bias, which is something that Spotify realized soon after they launched. When Spotify launched, they launched a collaborative filter. The reason they did that should be obvious. A collaborative filter, you can launch in a matter of weeks.

It's super easy. You don't have to understand music in depth. You don't have to understand movies in depth. You just say, "People who listen to this also listen to that. People who watch this also watch that." You can get these systems up and running in a few weeks, and they have a huge impact on purchases, on engagement, and so on. And so there's so much value, you just roll out the first system that you can. And then they started measuring, and they observed what we found in our research. And so then they said, "Okay, we now need to break the popularity bias of collaborative filters." And they wanted to then incorporate Pandora's approach. But Pandora's approach took years to build, and Pandora was not started as a commercial company. It came out of a research project called the Music Genome Project, so the initial R&D was not funded by for-profit investors.

So Spotify kind of had to figure out how to automate this. So what they did was they had a few musicians listen to a bunch of songs and rate these songs for things like syncopation and vamping, whatever those things mean. And then they said, "Okay, that's now the training data set for machine learning algorithms." So now the machine learning algorithm looks at the pattern in the small sample of songs that humans have listened to and rated. And once they've learned the patterns, they can look at all of the other songs in the library and automatically give them scores for how rhythmic they are, how much musical instrumentation there is, and so on. And of course, since they've done that—

Ravi: It's a hybrid system.

Kartik: --it's now a hybrid system. It gives you the benefit of both. There's a social appeal in consuming what others are consuming, and there is the value of also discovering new stuff. And their new system actually drives a lot of listens for-

Ravi: Niche artists.

Kartik: --niche artists. Exactly. And they say, for niche artists, almost 80% of the listens are coming from the algorithms rather than people coming in and searching for these niche artists.

Ravi: Fabulous. So maybe we'll switch over to some from the audience.

Audience Member: So this is an easy one. What is the next thing in AI that is definitely going to transform everybody's lives in the short term, like how Facebook did or the internet did?

Kartik: Oh my goodness, that's the easy one.

Audience Member: I'm joking. I gave you the hardest one first.

Kartik: Yes, it's super hard. What is the next thing in AI? Well, I'll pick two, three things that I think are interesting, right? I won't pick any sensational stuff, but stuff that I think is going to be revolutionary over the long term. I think driverless cars is certainly something I'm immensely excited about. The rate of improvement of this technology is amazing, and I think it's going to open up just a whole new world of-- an entirely new industry around driverless cars. But also all the productivity gains. Imagine you have a drive of an hour, and you get that hour back. So it's just going to be phenomenal. So that I think is one area. And again, it's all about the AI there. So that's one.

I think another one, not so much as an industry but an area of AI that will become extremely important is what we were talking about, the fact that really fairness, accountability, and transparency, that is going to be extremely important. So we've talked about fairness. We haven't talked about some other areas. So today some of the best-performing machine learning models are often the most opaque. Some people call them black-box models, like deep learning models and so on. It can give you highly accurate predictions, but they cannot explain why, exactly, the decisions the modeler-- their explanation would be of the form that, "I suggested this loan should be rejected because neuron 52 lit up and neuron 88 lit up which caused neuron 500 to light up and so on," and that doesn't mean anything. So there's a lot of interest in what's called post hoc interpretations.

Ravi: To explain black box.

Kartik: Yes, the black box model. And that's an AI thing. I think explainable AI is what it's called. That was one of the terms people brought up that's going to be a fascinating area, and a great area for the MSBA students to invest in.

Ravi: or the business.

Kartik: And, of course, the PhD students too. So I think the area is pretty interesting. I also think the third area that I find interesting is, if I look out there, every start-up out there these days is an AI start-up. Everyone calls whatever they do as AI, and if I look at where they're applying AI, it's not the most obvious areas. I'm actually getting interested in this idea of applying AI--to areas where- people claim expertise, and you kind of suspect it's vaporware, and huge industries are built on that. So, for example, the things I think of wealth management, venture capital, healthcare, Hollywood, all of these areas. Imagine Hollywood, billions of dollars invested. There are five people in the room who will kind of make these decisions. Have that project. It's $100 million dollars. This does not get made. We've all heard these stories like Raiders of the Lost

Ark, rejected by most studios and then finally got made and things like that. I think, again, data can have a huge impact. If you look at venture capital, again, individuals making these decisions, taking in big money. But most of them cannot repeat the performance a few years later, and so, again, AI and machine learning could be huge there.

Ravi: That's fascinating, actually. So I guess one of the things we have seen here, over the last five years, while working with both students and the companies, that there is so much low-hanging fruit, actually, right, in applying machine learning to problems. One of the most interesting problems that we saw a few years back was for a company whose marketing division went out and promised next-day delivery to their customers. It's a big, big company, and that completely shocked operating performance because they just sort of-- they were not logistically prepared to deal with that. And somebody in the company was smart enough to say, "You know what? Can we predict whether somebody is going to order two times in a day?" Right? If they place an order in the morning, are they also going to place an order in the afternoon? And if they do that, then we just hold the order, and we don't call the place twice. We call them once in the day, right? And that was billions of dollars-- not really. It was tens of millions of dollars worth of savings for this company. And every spring semester, we have examples of this sort. So my sense is that there is, cumulatively, a lot of inefficiency in decision-making that's happening at every stage. And I think that's actually the big one, in some sense. I think that will add up to freeing up capacity to do a lot of other things, I think.

Kartik: Yeah, I mean, I think that's a really interesting application, and I think one thing that-- I agree with what you're saying, but one thing I'll add to that, and to your last point that it can free up capacity, I think there's all this debate about AI and what's the role for humans? I think at least for a long time-- and by long, I just mean as long as we are all alive, right? Things are changing quite rapidly, so I never predict what happens two or three years from now. I think for a long time, a lot of it is going to be freeing up our capacity so that we can focus on things that matter.

That are more value-adding and so on. And one interesting example, I work with this company called Replicant. Basically, Replicant is a call center application. So what they do is human voice-- it sounds like a human you're talking to but they take care of the calls that come in. And pretty much all-- and they're producing amazing results for their clients. And mostly what we see when you look at the data is that they're taking a lot of the really routine calls. They still cannot handle 100% of the calls, but they're on or near the point where they can take probably 50% of the calls, and their goal is to get to 80% in the next three, four years. And the point is that there are so many routine calls that are easy for them to take. And also call center people are often so-- it's a high stress-

It's a horrible job. And so this can take care of a lot of that, so they can focus on the harder, more interesting problems to solve. And many other settings, so I think that's an interesting way to think about AI for the short-term medium.

And so it's finding those complementarities between human intelligence and machine intelligence, right? And what do we use which for, in some senses, right? That's kind of the holy grail. It's almost a molecular thing, usually with your own life as much as it is for the company or, I would say, for the nation, right, in terms of figuring out if you want to have an overall AI strategy, right?

And I think the idea of the AI strategy at a country level, then a company level and an individual level-- and in an individual level its, "What skills should I apply? What should I get changed so I don't get replaced by machines?" So at the company level, it's figuring out, "Where should I deploy AI? How should I reallocate the human resources?" And I know your colleague, Abhinav Gupta, is doing some very interesting work on what should be delegated to humans, what should be delegated to AI. I think that's a pretty interesting way to look at things as well.

Well, so we will get into that explainability aspect in a little bit, but I just want to make sure that we-- because that's an interesting aspect and an interesting passion already. But are there any other burning, interesting question? If anybody wants to speak out that question, you can raise your hands as well. Let's see if there's something there already.

Audience Member: So one of the questions-- as you mentioned, you talked a bit about fake news as an issue. What are some of your thoughts about how to deal with that issue going forward?

Kartik: Yeah. Fake news. It's an interesting one, but that's one case that I made the mistake of making a prediction. So I actually wrote an oped - I think it was earlier this year - that fake news will not be as important an issue in two or three years. And I'm getting very nervous about that prediction. But bottom line, I think there's a lot of really interesting research going on in terms of fake news detection, and in terms of these academic papers, you're seeing detection rates at 80, 90 percent. So really amazing results and performance, which is what led me to write an op-ed. At the time, I was promoting a book, some time back, and they said, "Give us something that is very sensational." And I said, "I don't know, people want to hear about fake news, I would think." Go op-ed, because I read a few of these papers around that time, and they were showing amazing results. But where it gets interesting or hard is-- of course, if you're getting really good at detecting fake news, the other side is not fixed. They're going to respond to that, and then it's like this whole cat and mouse game, right? And so that is what makes it hard.

But, I think, in terms of fake news detection the algorithms that are being built are actually doing a really good job of detecting fake news. These are actually being used in many settings. The

problem becomes in highly decentralized environments or in distributed environments. Think about WhatsApp. How do you detect fake news on WhatsApp? Because fundamentally, it's an encrypted platform, and Facebook, which owns WhatsApp, is not allowed to read what is in there. And so what do you do there? So I think there's a whole-these systems fit in a social system, and the social systems are complex. And that, I think, is why fake news is problematic. But if you give me a story, and say, "Can you detect if this is fake or not?" these algorithms can do pretty well already. And they'll get even better in the next two, three years. But how do we deal with decentralized systems? How do we deal with the fact that the other side is going to look in? The people working in this area are a little bit of-- should hopefully be busy for another 10, 20 years to cover that.

Ravi: There was a fascinating story about an Indian reporter actually who embedded herself in these WhatsApp groups that were-- so again, the last Indian election-- if you think the US election was kind of manipulated. Actually, the world's last election, which was in India very recently was very, very systematically, I guess, controlled, I would say, by the ruling party right now. And they had a machine. Literally, there was, every day-- so they could engineer trending topics on Twitter, at will, in some senses, right? So one of the interesting things-- I know it was that there was a lady who embedded herself in all these WhatsApp groups, and was kind of, after the fact, revealing the level of some of the vitriol and other things that they were putting out there. So there was basically no other way. There had to be some spy in there, a human spy, actually, to tackle that.

Kartik: I think that's an interesting one. I had a conversation with a friend who kept complaining about all the fake news and all the propaganda and so on. And I was kind of telling him something similar which is, "Stop complaining. If you care about this, find two other friends, each of you go join the IT cell of the three political parties. Sit there for three months, and then do an expose on what happened." But coming back to this point, I think anybody can create fake news. It's a technology problem but it's also an education problem. I think the technology problem is easier to solve. It's what I'm trying to do.

Ravi: Right. But the people, they actually want to hear the fake news, right? I mean, so it's really what's happening

Kartik: Yeah. It's the human side of things. There's people you will not bother to check. They will gladly circulate the fake news. They probably even suspect it's very fake but they won't bother. So there's all of those kinds of issues.

Audience Member: So in regards to training data, there's typically unconscious bias because of the individual who's providing that information has a specific worldview. So as it relates to the historical information that we're putting into this algorithm that has training data that is biased, what are the standards or different policies or whatnot that are being abided by in regards to

research? To think about the greater good and how we're not thinking about AI structures for historical data, though it's happened in the past, and making sure that as a society, we're continuing to move forward and continue to diversify the program?

Kartik: Yeah. So the short answer to your question, and it is an unfortunate answer, at least as of today, there are no standards for that. So there's so many people who have different ideas, but it isn't like there's a standard that everyone kind of adopts. And I do believe there's value in having a standard. And to say that I haven't been out there, I have done these checks. In fact, one of the things that's part of this book I've been advocating is that ML systems need to have an audit process and that audit process should include audits of the input, audits of the model, and audits of the output. And then within audits of the input, there should be some test for fairness. There should be a test for data quality. The source of the data for the model, again, is benchmarking against alternatives. So there's a bunch of these things that are prescribed. And there are a few other ideas out there that are for standards but I do believe that an audit process is needed for AI that has any social consequence. So loan approvals, recruiting, news curation, all of those, I think there should be an audit system. A company should actually be able to say that, "We actually did a third-party audit where we went, 'The audit looked at these things and improved the company.'" And an audit sounds like a dirty word for a lot of CEOs because it's another compliance headache, and I get it. But it could also mean just an internal QA process. What is interesting is that software typically has a QA process. Every company has test engineers who will do this.

Ravi: There's a whole industry around it.

Kartik: There's a whole industry around it. But data science does not have a QA process. You would never hear somebody say, "I've written this machine learning system. And then there's this whole QA process before it went into production. And in fact, even after it goes into production, it's not constantly learning from your data so there still needs to be continuous QA." There's no such thing for machine learning.

Ravi: And why is that?

Kartik: It's all new. I think we're just learning about these problems now. I think it has to happen. The question is what is the thing that pushes us there? I think of two or three potential ways in which industry helps a QA process for machine learning. One is that it actually impacts vendors. So you look at the software industry. A lot of the testing has it's origination in the ISO standards, ISO 9000, and so on, and all of these testings came from there. And you went out there as a vendor and you said, "My software is ISO certified technology with the same process." And so, maybe that is one way this happens that you have certification that helps you differentiate

yourself from a competitor. The other is regulation and that becomes another forcing function. But other than that, there is no forcing function as such.

Audience Member: So, you mentioned about trade-off between different areas objectives. Now, trade-off, I don't know, seems to also go above that versus-- there's some recent research that shows there's a trade-off between fairness and transparency, or explainability of the model. So how should we think about what's the trade-off to move with, when the trade-off exists between two desirable properties that we really want to have at the same time?

Kartik: Yeah, that's an interesting one. I don't know the answer to that, and I'm just being honest about it. It's a tough one. And I think it depends so much on the application context. There are some contexts where you would say fairness isn't the number one thing. We just want to know that we understand. Like if I'm using a way to make investments in the stock market, I'm mostly focused on, "Do I understand how these decisions were made--?"

Ravi: Accuracy, right?

Kartik: It's accuracy, and then there is also a trade-off between accuracy and the model interpretability or model transparency. And some of the more transparent models are not the most accurate. So what you do there is you take the most accurate model, and if you do what we call post-op interpretation-

Ravi: It's funny, but in finance-- I mean, if I can predict that simply following the up and down or even at 55%, that's already good numbers, right?

Kartik: It's interesting, actually. Many of them have said we haven't deployed, so-- what is that company, Ray Dalio's company?

So I met with somebody in Bridgewater, and he said that we didn't deploy-

Ravi: Because they couldn’t understand it.

Kartik: Because they couldn't understand it. Basically, we said, "Because you're making huge bets. If something goes wrong, the explanation cannot be that the internet decided this." So that's why they kind of said, "We need explanations, else we could not do this." And I think this is a whole new area that's going to be hot and I think, again, the trade-offs depend on the context. It's very hard to say this is a standard way to think of it. I'm actually part of a group-- it's a group of banks that has come together to figure out AI governance standards for banks. So I'm part of that group. And the conversations there are about what is important for the banks. And again, even within that, it varies based on the division of the bank. So with anti-money laundering, they're using ML. For credit fraud, they're using ML. They're using it for internal purposes. For one, they care about fairness. For another one, they care about performance

mostly. For something else, they care about explanations. And so, as part of that, we're trying to come up with a standard for banks. But it's very hard, because we can come up with something, then somebody else's says, "No, but in my division, we don't care about this and that." So I would say, "Okay, I can say write this portion."

Ravi: Well, I think that's fabulous. So I guess maybe one final question. So what's the next book?

Kartik: What's the next book?... a Podcast.