**CRAIG:** Hi, I'm Craig Smith. And this is I on AI this week. I speak to Michael Kearns, a computer scientist professor and national center chair at the university of Pennsylvania and the founding director of pens sing program in networked and social systems. Michael is also an Amazon scholar and has been working at Amazon on privacy and fairness.

**CRAIG:** We talked about differential privacy, how Amazon's research approach differs from its peers and how AI will eventually permeate all aspects of our lives. Before we begin. Let's take a moment to thank our sponsor ClearML an open-source ML ops solution. You can give them a try at clear.ml.

**CRAIG:** Tell them Eye on AI sent you. Now here's Michael Kearns. I hope you find the conversation as enlightening as I did.

**CRAIG:** Michael, it's great to talk. I generally start by having people introduce themselves and give a little bit of background where you got your education. What you were doing before you came to Amazon and what you're doing at Amazon now, and then I have some questions.

**KEARNS:** Okay, great.

**KEARNS:** So, the short version is I am a California native. I grew up in San Diego from an academic family, went to undergraduate at Berkeley, and then got a PhD in computer science from Harvard, and then more or less spent the nineties at the late great Bell Laboratories in basic research, which was a great era for early machine learning work then. Had terrific colleagues. And then I moved to the University of Pennsylvania faculty in 2002. And I've been there ever since and still am. And I joined Amazon just a little over two years ago as part of their very innovative Amazon scholar program, which is meant to let academics like me spend very serious time working with Amazonians on various technical problems in our areas of research interest while still having our academic life as well. And topic wise, I am a lifelong card-carrying machine learning researcher.

**KEARNS:** So, I started working in machine learning in the late eighties. Which at that time was a subfield of the then discredited larger field of artificial intelligence. And so, suffice to say, I've seen a great deal of change over the past few decades.

**KEARNS:** I've always had a fairly rich and diverse extracurricular life outside of academia with consulting and the like, and so in particular, I've spent a long time doing quantitative trading work on wall street, which I stopped in order to join Amazon a couple of years ago.

**KEARNS:** And so, when the deep learning revolution happened early last decade, it seems so long ago now, I was in industry, my alter ego was on wall street where I still don't think deep learning has made the inroads it has in other applications areas like computer vision and speech and things like that.

**CRAIG:** Even with the advent of transformers and the rise of sequence-to-sequence learning seems like that would be perfect for wall street.

**KEARNS:** the most promising area for deep learning on wall street is probably in the area of high frequency trading, because that's the area where you have the volumes and the density and granularity of data.

**KEARNS:** The data that's used in high frequency trading is millisecond or even microsecond timescale data about not just the trades that get executed, but all the bids and asks and people entering and withdrawing orders and things like that.

**KEARNS:** I like to follow the markets. I find the markets interesting. But I kept my activity to commercial activity My personal investments are fairly mundane as you not surprisingly find with a lot of people on wall street, especially the quant types.

**KEARNS:** They're not the type to be stock pickers as they're called.

**CRAIG:** I had an opportunity to test drive a no code platform called Akkio, and they had a fairly standard classification algorithm and it's a drag and drop interface. And I got a guy involved in horse racing to feed me data and it did surprisingly well in picking the favorites and every now and then it would have an outlier then, which, if it hit, paid a lot of money.

**KEARNS:** Did you pick rich strike for the Kentucky Derby this year?

**CRAIG:** No, actually this is last year, but then my data guy pulled out, so I haven't done it, but I've always been fascinated with using deep learning. Maybe since I'm a journalist and never made any money.

**CRAIG:** Tell me a little bit about what you're doing. Is most of your work right now at Amazon or is it at the university?

**KEARNS:** It's a pretty healthy mixture. The scholar program is very flexible. So, this is actually the third summer that I'm full time at Amazon, which, doesn't mean that I completely drop all of my academic activities. It's less during the academic year when I'm teaching, there's just more activity going on with graduate students and recruiting, graduate students and grant writing and things like that.

**KEARNS:** But I would say it's been, all in roughly an even balance over the past couple of years.

**CRAIG:** And you're working primarily on privacy and safety at Amazon.

**KEARNS:** Yeah. So, I was brought in to help think about responsible AI, generally speaking. So, the topics you mentioned, also things like fairness in machine learning, making sure that trained models are not biased or have worse performance on one demographic group versus another demographic group. And so, I work with everybody from scientists to engineers, to developers, to people in the product and service teams, thinking about both the technical aspects of this, and also, what are good definitions of fairness?

**KEARNS:** What are good definitions of privacy? Those are very big words. But at the end of the day, if you're going to audit for a privacy or fairness property, you need to be precise about what you're auditing for. And then if you find something that you think needs fixing, then you have to decide what to do about it.

**KEARNS:** And in the case of fairness, it could be simply that you need more data on some group or type of data that you're performing less well on. That can actually be quite time consuming because if you don't have the data, it probably means it's not easy to acquire. And so, you might have to go out and actually curate and annotate it. A lot of my work is technical in nature, but a lot of it is what you might call more conceptual, which is what should we even be looking for.

**CRAIG:** Guiding the data scientists on what to focus on.

**KEARNS:** Exactly. Yeah.

**CRAIG:** I'm curious about work at Amazon. Amazon developed a little differently in the field of AI than say Google or Microsoft.

**CRAIG:** It was much more product focused from the beginning. It hasn't spent a lot of time, from what I can see, on expanding ' the toolbox of ideas,' as a guy I interviewed recently talked about. it's more concrete in its work. Is that right? Yeah. I have a particular perspective on this.

**KEARNS:** So, Amazon is definitely structured differently in the way it approaches science and research than its peers. If you go back to early in my career at bell laboratories, many people hold bell laboratories up as the pinnacle of industrial research and this changed over time. But even when I was there, research was a very separate organization and research reported directly into the C-suite.

**KEARNS:** So, it was not the case that for instance, research every year had to justify its existence, its budget, et cetera, to particular business units and the like, and Microsoft in particular followed that route in designing their basic research. Google less so at the beginning.

**KEARNS:** But even though I experienced bell labs. I like the way that Amazon is structured very much, because there are a lot of great scientists and researchers, they generally sit in or very near to product teams. But in general, they report to science leadership.

**KEARNS:** And so, it's great for somebody like me, because if I was looking for a pure research environment, I would just lever leave my office at Penn. And Amazon is the right environment for people who are looking to apply their research and also get research ideas from practical problems. And by having people really embedded, close to product teams, that works very well.

**KEARNS:** In particular it's been a great experience for me because I've been able to really see what product teams do, how they work, how ideas get germinated and then disseminated throughout the organization.

**CRAIG:** Can you talk a little bit about that? Because I was talking to Rohit Prasad earlier and he was saying how Amazon is very much bottom up in its research that it's driven by the market.

**CRAIG:** And how does that affect your work? And then talk a little bit about which product group you're closest to. I've never been in one of these organizations to watch the work, how it happens.

**KEARNS:** Yeah. So, remembering that I have limited points of comparison since my prior industry experience was mainly on wall street, which is a very different beast entirely.

**KEARNS:** But especially AWS has somewhat of a reputation for Land of many startups within a very large organization. And there's a lot of truth to that. So, in my experience, it's pretty easy for small groups of people to get an idea for a new product or service. And there's a very rational process for disseminating that idea and socializing it and getting feedback on it and iterating on it and modifying it.

**KEARNS:** And then eventually getting the resources to go out and build it. So, in the short time I've been there, I've been part of that process for some privacy related work that we're doing. And also, just more generally within the machine learning space, including in areas like fairness, explainability and the like. Again, my data is limited, but it does feel like there's a minimum of bureaucracy.

**KEARNS:** It's a very flat organization. A lot of the great ideas come about organically. It's not, some invisible voice from on high saying, thou shall build this. It's really, the ideas are coming from the researchers, the engineers, the developers, and of course the leadership as well.

**CRAIG:** And, when you're working on privacy, for example there are a standard set of strategies for achieving privacy. Are you working on new algorithms or are you working on architectures? Are you working on web three? For example?

**KEARNS:** Yeah. So, the field of privacy has been and is continuing to undergo a bit of a sea change. So let me start with the origins of that sea change from the science side.

**KEARNS:** If you go back 20 years, by far and away, the most predominant notion of data privacy was anonymization techniques. So, you take a dataset and, there is personally identifying information in it. And so, you perhaps eradicate entire columns, social security numbers and the like, things that are unique identifiers for a person, but then you might correctly surmise that if I know enough, apparently innocuous facts about you, like your age, where you live, whether you use a Mac or a PC, et cetera, that might also act as a unique fingerprint for you.

**KEARNS:** And so, the idea of anonymization is that somehow you do enough redaction and maybe coarsening - instead of getting your exact age, I put it into 10-year bins or something like this. The idea is that somehow it should be difficult to reverse engineer your identity or your data. And here, I will bookmark that I'm expressing a scientific opinion, but basically in the first decade of this century, people gradually realized that anonymization techniques are extremely fragile and brittle.

**KEARNS:** Not that they're not useful in many circumstances. But in particular, there aren't any formal guarantees you can give about anonymization techniques for the most part. And at the same time that was happening, this alternative notion of privacy, which is called differential privacy, which sort of begins with a definition that provides very strong, formal guarantees, was proposed in something like 2006.

**KEARNS:** And essentially over the last 15 years, scientifically, that field has developed greatly. And so now you're starting to see deployments of it. And at a high level, the idea behind differential privacy is to add noise to computations in a way that on the one hand preserves its usefulness, in the sense that aggregate statistics for example are approximately preserved.

**KEARNS:** But the addition of the randomness to the computation obscures the contribution of any individual person's data in a very precise, formal way. And as an example of the current moonshot for differential privacy, as you may know, at some point several years ago, the census bureau in the United States decided that every single statistic released from the 2020 us census would be done under the constraint of differential privacy. This is not to say that this is the only notion of privacy, but it's definitely, the one that has the sort of, if you're really concerned about having strong, clear guarantees about what you can say to an individual regarding privacy of their data, the right answer right now.

**CRAIG:** Differential privacy you're talking about, is that making it into Amazon products?

**KEARNS:** I guess I shouldn't comment on product and product releases, but there's a largish group of us who are working hard on it and thinking about it and working on the science aspects of it.

**CRAIG:** And when you say the science aspects of it, you're talking about specific models that you build.

**KEARNS:** Yeah. Or it might be training models in a differentially private fashion, which is a subtopic in the field. And so let me parse that a little bit. In my mind. And in many people's minds, privacy and security are complimentary technology.

**KEARNS:** So, what I mean by security, which writ large would be the domain of cryptography and related areas. The problem there is you've got sensitive data. You want to keep it locked down and control access to it. You maybe have a sensitive data set, maybe it's medical records and you want to make some good societal use of that database of medical records.

**KEARNS:** You want to use it to build a predictive model for a rare disease. So, you train, let's say a neural network on that data set and it works well. And now you'd like to share your findings with the world. You want to publish a paper in Nature, describing your model. Maybe you even want to give the model to other groups to use or other hospitals to use.

**KEARNS:** If you don't make a special effort to prevent it, you should expect that trained model might leak or exfiltrate the training data so that, somebody looking just at the parameters of the neural network might be able to reverse engineer data points that were in the training data. So, the idea behind differentially private model training is to train the model in a way that as I mentioned, adds some noise to the process of training the model in a way that guarantees that, that can't happen.

**KEARNS:** And so, this would be one use case. Another use case is the mouthful of differentially, private synthetic data generation, which is another topic that has like a literature within privacy. And the idea there is rather than training a model or answering a query for you on the dataset in a differentially private way, I actually just want to give you a synthetic dataset that looks like the original but has these very strong privacy guarantees.

**KEARNS:** So, it'll look like the original and if it was medical records before the dataset you'll be given will look like synthetic medical records, but there will be a promise that nothing you can do with that dataset could actually help you reverse engineer the real medical records in the original.

**KEARNS:** But on the other hand, this dataset would be useful for you to, for instance, ask about the correlation between this symptom and this particular condition, for instance. So, it's again, preserving high level aggregate statistics while providing these very strong individual protections.

**CRAIG:** And so, both of those you're working at the data prep end.

**KEARNS:** Yeah. The data prep end or the model training end. Yeah. This is all very core underlying technology. Now, what exactly you would do to turn this into a product is left to your imagination. But there's something now called open DP, which is an academic and industry collaboration, which is an open-source toolkit for differential privacy that's out there and people use.

**CRAIG:** When you're working at Amazon beyond the conceptual part of the job, what kind of resources are available and is that important to your work that would not be available sitting at the university?

**KEARNS:** My gut reaction to that question is that the first resource that's very valuable and different than what I have in academia is the makeup of the people that I work with.

**KEARNS:** They're from a very different background and experience. They have a very different worldview, even on the same topics that I'm interested in as an academic. And I might basically describe that worldview as more applied. You'll propose an idea that seems like a good one on paper or from a scientific standpoint.

**KEARNS:** And then people will start asking hard questions about how you would really make that work. And, whether you'll be able to explain it to customers and whether it'll scale well, et cetera. So, while it is also the case that for instance, the compute power available at a place like AWS is something that I certainly don't have lying around.

**KEARNS:** In my academic life. That's an important resource, but by the time you get to the point that you're using that resource it's sign, that you've decided that there's something interesting to try and do. And so, for me, what's more important is what came before that, which was the sort of social process to decide.

**KEARNS:** Now we think that this might be an interesting experiment. If we could get this to work at scale on very large data sets. For instance, we think that there might be, an interesting product or service offering there.

**CRAIG:** And in the academic world, if you reach the point that you thought it was worth, there are ways to get the compute, for example.

**KEARNS:** Yeah. I could always become an AWS customer and do it that way as well. the real difference is just the nature of collaboration. Academia is, in many ways, a very formal, structured place, more so than industry for the most part. And at Penn, I have a group of graduate students that I, I essentially run with a close faculty colleague who's also an Amazon scholar named Aaron Roth. But we have our group of graduate students and postdocs, and we do research with them. There's not this sort of free form socialization process, that there is like a place like Amazon, where you have an idea and it's not just scientists commenting on it, but it's, user interface, designers, its product people, it's product managers.

**KEARNS:** It is software developers and engineers. There just wouldn't be the environment in academia for me to say oh, hey, that this is a practical idea that might actually be worth thinking about as a commercial product or service. There just wouldn't be the right mixture of people around to critique it the way there is at Amazon.

**CRAIG:** How much of what you're working on at Amazon. And then in your academic life is deep learning. Are you still involved in other forms of artificial intelligence?

**KEARNS:** Yeah. And in fact, I am not a deep learning specialist. And by that, I mean, the people who are real practitioners and engineers of it.

**KEARNS:** More what you might call a machine learning generalist, meaning, I know the field I've been around a long time. I have a broad view of the different subfields of it. Deep learning in particular is you're probably aware, has been, especially successful in what I would call the modalities.

**KEARNS:** So, it's, things like speech processing or computer vision, where the data is very voluminous. It's very dense. The data has some structure to it. What do I mean by that? So, by structure or I might even use the fancy word topology, which would be a little bit more accurate, but what I mean by that word is if you look at a speech signal, look at the actual wave form of the speech signal. It's a pretty good bet that right nearby parts of the signal are related to each other. They might be part of the same phoneme. They might be a break between phoneme, things like that. And so, there's this sequential structure to the data, same thing in vision, right?

**KEARNS:** If I give you a digital image, there's a good chance that two nearby pixels are related to each. They might be part of the same object they might be on adjacent objects, which is important because then there's an edge or boundary between them. In contrast, many machine learning problems are what people would call tabular data, meaning it's just like I have a data set and one column is age and another column is, employer.

**KEARNS:** And another column is, where you went to school and maybe this is, a personnel database, for example, each column is representing a completely different thing and deep learning has been especially successful in the first type of data.

**KEARNS:** And it's not that I'm a specialist in tabular data, but it's, especially the computer vision, speech and similar communities that are really doing very hands on applied work to solve problems in those subfields.

**CRAIG:** Do you have any opinions about where it's going? For example, right now, there's this, seems to me, over emphasis on the promise of large language models and there's less emphasis on novel algorithms and how all of this is moving toward a more synthesized artificial intelligence that could approximate human intelligence.

**KEARNS:** Yeah. Large language models again are an instance of the successes of deep learning.

**KEARNS:** I do think deep learning still has some head room, right? These large language models are certainly interesting, but they also highlight some of the shortcomings of modern AI that have been the shortcomings of AI for many decades.

**KEARNS:** This includes the fact that they're not good at logical reasoning. They don't have common sense. In this talk I saw the other day, you could lead one of these models in a conversation to say very different things at the end, actually contradictory things, depending on the actual sequence you took it through.

**KEARNS:** There's a long sequence of interaction between the researcher and the large language model and long story short at the end of one of them, the model concludes only female ducks can lay eggs. And at the other one, it's like all ducks can lay eggs and it's quite insistent about it. And both of these transcripts are entirely rational in their self-contained way, and one of them reaches a true conclusion and the other one reaches an untrue conclusion. So, on the one hand. It's incredibly impressive. If you had shown people, something like this, back in the 1990s, their heads would've fallen off. Just the fact that there's this very rational, cogent conversation going on between a person and a machine.

**KEARNS:** But you put these two transcripts side by side and you realize, okay, there's something not quite right here. There's some kind of very sophisticated mimicry going on, and this is to be expected, right? There's no mystery that these things aren't synthesizing from whole cloth things that aren't at least implicitly or even literally contained in the extremely large data sets that they were trained on.

**CRAIG:** That's very interesting because from the public's point of view, it's indistinguishable, the, as sentience versus mimicry.

**CRAIG:** And then it may be human consciousness and reasoning is mimicry too, but just extremely complex and evolved.

**KEARNS:** Obviously I cannot refute that sitting here. It's not an implausible idea, it could be that our own intelligence is nothing but a matter of extremely sophisticated mimicry.

**CRAIG:** On the privacy topic, there's been a lot made about the advent of quantum computing and how that will blow apart the privacy paradigms that we've developed. Is that something that you think about or have looked at?

**KEARNS:** I'm not a quantum specialist. Again, this is an area with a long history and a lot of very smart people working in it.

**KEARNS:** I Was trained as a theoretician, so I especially follow the theoretical developments. And what I mean by that is in the same way that there are formal mathematical models of what classical digital computers can do. There is, a version for quantum computing as well.

**KEARNS:** And a lot of the theory work tries to identify whether from a formal mathematical standpoint, there are certain computational problems that can be solved exponentially faster than they could classically. But we're not close to wide scale deployed quantum computers that outstrip classical digital computation. But if that happened, it would be a big deal. When you step back from the hype and look at the science, there's something deep to think about here. That if it were true, would fundamentally alter our thoughts about computation, and it would have some real implications, not all of them positive. So, as you probably know if large scale quantum computation becomes a reality, then essentially all of the encryption schemes we rely on for security right now would essentially be compromised.

**CRAIG:** And in terms of applied AI whether it's good old-fashioned AI or deep learning, I've been talking to people about ambient intelligence and the spread of these cheap sensors and synthesizing data from them.

**CRAIG:** I know you don't want to predict the future, but do you have **a vision for how AI will be** in in our lives in the next couple of years?

**KEARNS:** so clearly to the extent that it's possible for us to systematically gather new sources of valuable data that we don't currently have, there will be all kinds of great applications of AI and machine learning, and I've learned about many of them by just hanging around at Amazon.

**KEARNS:** So, there's a line of products within AWS that are meant for monitoring industrial equipment. So, you put like a sensor on a machine, and it gathers data and understands the vibrational patterns of the normal operation of that machine, and then uses machine learning and time series analysis to detect when suddenly there's some knocking sound occurring at some frequency that's not normal. And then the model says somebody should go check out this unit. Things like that are great. And that's an example, of course, where you're gathering data that doesn't have to do with people's social activity and the like.

**KEARNS:** Obviously in the last 10 years or so, the power of machine learning for making very useful predictions about individuals and recommendations to those individuals like, navigation apps, and so a lot of the research I've been doing in the last seven years, and many others doing it with me and what's now become responsible AI is trying to think hard about, okay, this is great.

**KEARNS:** But it's not always good for society. There can be side effects that we didn't intend. And if you dug into a lot of the underlying science in this area, you'd realize that a lot of times the problem isn't because somebody did something wrong or was lazy or was biased in some way.

**KEARNS:** It's just that machine learning never gives you something that you didn't explicitly ask for. So, if you tell it, minimize the overall error across the population, that's what it's going to do. And when I said minimize the error across the entire population, I didn't say, oh, and by the way, could you also make sure that you equalize the error rate between men and women?

**KEARNS:** Okay. That's a different statement. It seems related, but it's different. The problems can arise when, the model that minimized the overall error does so at the expense of not doing as well on some other group, but then when you realize what the problem is, there are things you can do about it.

**KEARNS:** And one of them has changed the way you're training. One of them is the way you curate data or how you collect it. And I'm not claiming this will solve all of the unintended side effects of AI, but it's a good start at least on the science end of things.

**CRAIG:** These unintended consequences that you're talking about are pretty immediate in the outputs, but you can't necessarily see how the outputs are going to have knock on effects. Do you think that we’ll keep discovering negative effects

**KEARNS:** At a high level. Yeah. Because you can't anticipate everything. But of course, the most important thing to do is address the highest order ones first the biggest, most obvious ones. And then if you get to the point where yes, there are side effects that we didn't intend and don't want, but they're getting smaller and smaller that's progress.

**CRAIG:** That's it for this week's episode of Eye on AI, if you want a transcript of the conversation as always, you can find one on our website eye-on.ai. I want to thank Michael for his time, and I want to thank ClearML for their continued support of the podcast. If you're looking for an MLOps platform, go visit clear.ml.

**CRAIG:** And remember, the singularity may not be near, but AI is about to change your world. So, pay attention.