PROF AI PODCAST

AUDIO TRANSCRIPT

FERNANDO LUCINI: So, I'm Fernando Lucini. I'm the Group Lead for Data Science and Human Ingenuity for Accenture. And with me today is Zachery Anderson, Chief Data and Analytics Officer from NatWest Group. Hi, Zach.

ZACHERY ANDERSON: Hey, how you are doing?

FERNANDO LUCINI: I am good. We just traveled to a little bit of technology difficulty to get together, so we're saying hello for the second time.

Today, I think, Zach and I wanted to have a chat around the personalization of AI and I think – and, Zach, remind me if this is – I think it started at one point with, oh with my obsession with personalization and then jumping you in a meeting around whether you felt the same way. Maybe it's worth, Zach, talking for a couple of minutes about of our interpretations of personalization as I think that's probably a fun way to begin and let me give the a one-minute version. But my background is in software engineering, so nobody's felt for me like even though software was not professionalized in the sense of having like government certifications and all this. But the good thing is it felt like along the period of 10 years or 15 years, it informally professionalized in that we all knew what our role was, testing for development and building versus this or that. And then the software world as in the product world marched along and made us all organized very nicely together. And it felt to me like the artificial intelligence world needed to start functioning in the same way.

And, again, just to kind of jump Zach in a meeting to say, Zach, do you feel the same? And thankfully, you seemed to align with that.

ZACHERY ANDERSON: It's interesting because I come at it from the other side, right, which is I'm really an analyst by training, not a software engineer. And I learned software and coding as a byproduct of being an analyst. But coming up through that path and started as an analyst long before anybody knew what a data scientist was and kind of went through the path of that going from, you know, I've been called in my career analyst, advanced analyst, statistician, engineer – I mean I think I've had economist, macroeconomist, microeconomist, I've had lots of different titles over time in my career. But I've come to the same conclusion as you, as I think, increasingly, what we're doing is building software and so, we need to adopt software standards in order to do good data science or AI or ML broadly in order to be able to do that.

And I also think the things we're doing, when we were statisticians on the analyst side, what we did was receive data, make model and handoff and that's just not what we're doing anymore when we're building models. We're building model pipelines. We care deeply about those pipelines, how performant they are, what the quality of them is, checking all of those things. And so, we aren't doing just simple statistical model building anymore. We're building code and software that needs to be treated as such.

So that's how I came to my conclusion of it. It's kind of from the opposite side, but same issue, right.

FERNANDO LUCINI: Absolutely. And it'd be fascinating to talk a little bit about the lessons that you put in place in NatWest Group as well about how you've tried to professionalize that as an example of landing and working in a big organization of how to do that. Obviously, I come at it from an Accenture perspective as in trying to serve customers and trying to professionalize them and get them to the places and interesting conversation.

And I think the first sort of mental model that I have is right, so if you land in our organization and you already have a lot of people, because let's not forget, in the last two or three years, people haven't been sleeping, right. We have people in places. We have data scientists. We have engineers. We have all these things. So what in your mind do you think is – let's think about the roles? What is the bad boys, do you think is like the roles that you would need to professionalize – you know, if you and I were talking to another one of our peers in the industry, what is about your mine thinking?

ZACHERY ANDERSON: Well, I mean the roles – I think in general what we're moving from, especially in the analyst community is kind of generalists, people that are analysts roughly know how to get their own data based on SQL, can maybe run some models or in the process of learning how to run models, to people who are specialized. So to data engineers, to ML Ops people that are actually running models and so, data engineering, sourcing data and ML Ops on the other side managing a model after it's built.

And I think the thing that's interesting though is although I think in specialization, I actually want people to have more broad skills. I think the shift from what we were talking about before, which is statisticians who receive data and then, did a regression and then output a mathematical model that was used in a scoring system is pretty different than what we have now in terms of the tools that we're using, in terms of the code that we're using, the complexity of it, how you need to debug it and test it and understand it.

And so, I like my – we're looking for people that can do many of those things and crossover. They don't have to have – I actually don't like people who are really super specialized to a certain area. I like people to be able to crossover. I want data – we talk about data scientists as a job category at the bank. We want data scientists to go get their own data. We want them to be able to write their own code. We want them to be operating with GitHub. We want them to be able to do monitoring of their models.

And so, I'm looking for people that can crossover now. In a project team, you might specialize a little bit more, but the key to this, I think, and the key is good software development is a holistic view of the process from first co-creation to production. And good software engineers and the training actually understand that whole process usually. I mean who's the most valued engineer now, full stack engineers that have gone the full way and I think that's where we're going with ML engineers and kind of the new titles is people that can crossover or cross a bunch of these things, but in a very technical way, not a kind of older school statistician where analysts way. But I say often, that the whole space that we're in of analytics has shifted from some IT data people and some statisticians and some analysts to becoming more and more sophisticated all the way up the tree. Because the tools are enabling it, the education's enabling it and so, it's a giant shift from generalist analyst to generalist software development, including modeling and analytics.

FERNANDO LUCINI: Right. I'll tell you what, I have a slightly different mental model. And probably I see compatible aside definitely. Maybe it's because I have the software engineering background and I came to science on the journey, which is I always think about data science is data science, right. and then, I think of the performance engineering and the skill ability engineering of the science. So I have my ML engineering thing on one side and like you, I do think that's a continuum, but it does feel - I was having a very interesting conversation with a friend from one of the good - the amazing U.S. universities about this. And we were sharing the perspective that more and more, the data science skills are more and more focused towards industry in a way that it is super specific.

So less and less, and this is, again, an argument we have, less and less do you have the need for a super specialist data scientist that doesn't understand in a gas and oil context of valve work or subsurface of topic. And that the tools for the rest of the stuff, for the performance engineering, for all that kind of stuff, the tools are getting better and better, so your knowledge can be broader and broader because it can be sustained broader and broader.

So at which point do you – you still that need that continuum of on one side, the super, super, super specific data scientist that understands the models, the life models of trading and loss and stuff like that, whilst you have a much broader ability to get people that have the right skills in performance and in scaling because of tools learned it and some degree of knowledge of data science.

My simple mind goes back to the fact of designs and performance.

ZACHERY ANDERSON: Yeah, and this is an engineering university or college.

FERNANDO LUCINI: This is a data science and machine learning department. It's interesting.

ZACHERY ANDERSON: Yeah, it is. I think we're still on that journey. I don't think the tools are that good yet, to be honest. I think performance and scaling still require some pretty good engineering tasks. So it's depending on what you're doing. And especially, when you have to connect into other systems, which is what we always have to do when we build a model.

And I think the tool, while the tools are advancing, the most advanced models are also continuing to evolve and outpace the tools in some ways and we have to continue to evolve that.

So I mean my sense is that there's – I think that part of that I agree with. I don't think the data scientists – like I really don't like the idea of people that just build models and can't get data and build pipelines. But I think the real challenge that I have with that idea of kind of data scientist and then somebody who scales, that's still a pretty big hand-off. Like I think we want to be producing not in production initial, but I think we want to be building production ready work very early in the process. I think of it much more – there's a model and actually we still use it in NatWest Labs, where we run modeling labs and that's a really popular model around the world in terms of adopting ML.

So it's an experiment. We kind of go find use cases and build it. But I think we're actually evolving towards factory, not a lab metaphor. And what I want my teams doing is building for production when they start a process. So we're not rebuilding pipelines and rebuilding these things and we need to be building – and to do that, you have to have more engineering skills earlier on and you have to have the right tools at play. Because there's this problem right now that, I think, the tools to build a model are more available than the tools to scale it. And as a result, now remember, you and I might be coming at it at two different angles and seeing the world differently, but I think it's way easier to build a model than it is to make it performant. And that step takes lots of time.

FERNANDO LUCINI: Or even getting the data, right?

ZACHERY ANDERSON: Yeah, or getting the data. Those are the steps. I'm trying to get us from idea to productionalized model injected into a decision-making system as fast as I possibly can. And right now, because all those things are divided up because people can't get access to data and because they can produce a model that it's not production ready, it has to be rewritten. All of those kinds of things 'cause the code hasn't been annotated to be handed off and get through governance. All of that kind of stuff just slows us down.

So I think right now what we're aiming at or what I'm thinking about at the bank is how do I turn us from a lab concept to a factory as fast as we can, so that we remove those big steps and we can get models into decision-making much faster.

FERNANDO LUCINI: I totally agree. And I wrote a piece on the death of the POC, which I've got a lot of slack for. But the truth is what I was trying to explain was not that we shouldn't kill the POC, of course, but the whole point was exactly that, which is how do set ourselves up for when you start any experiment you started is actually a production, almost prepared, it's a candidate for production. And it might not make it, but it's a candidate. You're not working it as an analyst, you just proof of this thing – that this particular model will work knowing that it can never make it into production. So I'm 100% with you.

ZACHERY ANDERSON: You know, it's interesting because I think you're exactly right, but let me – I think the idea that somewhere along the line proof of concept got kind of ruined. Instead of I'm going to prove that I can do the hardest bit of this, before I take on the rest. It became, I'm going to do something on production and not scalable in order to do it. But if scalability is the hard thing, that's what you should deal with first not last.

And so, I think that's the mindset is, everything you do should be planned to go into production, unless it's a true research project. So I'm good that you have no plan to do and it's for the education of the team or the education of the company or something. But assuming that you ever hope to have it in production, you should start with the assumption that it will be in production.

FERNANDO LUCINI: I have a theory about that, Zach, which is to some degree, some of this started. I mean, obviously, data science in machine learning and how it integrates to production and all that, but at the beginning of it is pretty much a research type endeavor. I've written software for decades and I start and I finish. I mean it's a foregone conclusion with some exceptions. But this was a research kind of endeavor where people did need to play around and things didn't work. And out of 10, 3 made it, 7 didn't make it. And in that, I think the POC did get a bad wrap or we created a motion that was not very helpful where we did the 3 to the 10, maybe that became 9 to the 1, with 9 experiments didn't go anywhere and then, we didn't think about integrating it to the company, dah, dah, dah, dah.

In that vein, let me ask you this. This is what fascinates me. Where would you choose to put standards and what standards would you apply? So if you had to think of standards, I do this exercise every now and then to myself and to customers. So if you have to think of standards in the flow of going from an experiment to finding the data, all the flows that go all the way down to actually pressing a button and putting it integrated into one of the bank systems, where would you be tempted to put in the standards, so both standards with you choose?

ZACHERY ANDERSON: I mean we've talked about some of them. I think they're around documentation, ability to share, ability for other people to pick it up. I definitely think data pipelines are important and there's lots of examples of people not thinking about their data pipelines as production. But then struggling at the end of the process to put the model into production because their selection of sample of their data pipeline was biased and they picked up on it and they picked up an incomplete sample and then, the model that they build is junk.

So I think if you know you're going to build a millisecond – or two millisecond return model because of where it's sitting, you better use data that conforms to that from the beginning or it's not – I don't care how great your model is, it's not never going to work 'cause scale is a key contributor to the delivery of the outcome. So it's really around – I mean if I was going to do it one by one, it would be around what – again, I kind of go back to what's the hard thing in this case? We know we have a two millisecond response that we need, so we better build the data pipeline to respond to that, any data we can't get in that pipeline and then when we can't conform to it isn't going to be part of that model.

So kind of go through the standardization process based on the use case that you're going to build. If it's s score and one time for analysis, then it's a different thing. But I think the biggest things that I have are end-to-end debug, I think completeness and robustness of the data pipeline and then, assume reuse and adaptation and storage and sharing of the code.

Those are the things that I think build something that becomes an asset for the firm even if you fail in the process.

FERNANDO LUCINI: Right. I'm with you. Because I think of it, I come from a different angle, I tend to think about what do you do in your company or government or whatever. There's partially that you'll do very well and just an example of this is code control. I mean how many times have you and I seen code controlling data science? Sometimes none. None. And I always use the joke, I don't know if you know this one, you know, the old put five software engineers in a room for five hours and after five hours, they everything done in terms of working together, code, controls, cycles, how you compile and all that stuff, testing regimes. You put five data scientists in a room with five hours and you'll get five amazing models done in five different languages and they would not always code.

ZACHERY ANDERSON: I think that's exactly true. And I didn't really understand, to be honest, I didn't really understand code control until I got to EA. When your company builds really complex software and things that are more complex than some of the Windows applications, you start to understand code control. When you have real compile issues because of the size of the code and the number and you have a thousand engineers working on a product, IO and code control becomes of itself.

FERNANDO LUCINI: It becomes of your final testing.

ZACHERY ANDERSON: But you're right. Before that, as a data scientist, growing up in the field, I didn't learn any of that. I learned that as a byproduct of being at EA and being in a real software development community.

FERNANDO LUCINI: And that's how I think of this. If you think of what a company does well, if code control is something you do very well, then don't put standards around that because you seem to have the ethos, you seem to have the muscle and there seems to be certainly in this world that you and I live in, there seems to be some places which are very common, data and the use of data seems to be a classic one you talked about. Super classic one.

As I say code control and version controlling of models is another classic one, model management, model classification, reusability of models seems to be another classic. So there's a few places there where I always look back and I think, well, what's stopping you from going into production? You think about it in simple ways and then, you work your way back as one level.

The second level is if you think about second or third or the consequences, you might be able to get away with no source control for an experiment, but once you're going into production, you can't get away with it. So one of the things that you can't get away with and what you go back –

ZACHERY ANDERSON: Yeah, and then you go a different direction.

FERNANDO LUCINI: And depending on what your company kind of does well and what it doesn't do so well, what bad habits come along the way. Then you can sort of unplug the standards question.

And then, every now and again, I'll find some company that has standards for everything. I have a famous example of a customer where we – I counted 45, 46 stage gates before we actually go to write one line of code. Let's not go down there either.

ZACHERY ANDERSON: That's why you have – I mean everything is describe and then constraints. And I think I said it upfront, my goal is to get to code fast – I mean get to production fast and do these other things, like have reusability, have quality data flows. And so, the standards should be conforming to both of those things. They should be not so onerous that you slow down the process brutally and make the team inefficient and at the same time, balance that with effective controls and both risk management and reusability.

FERNANDO LUCINI: Totally. Let me throw this one at you because I found it fascinating. Last week, I was talking to another – I was writing a piece, but I'll send you on another topic, but in writing that – I'm not trying to bait it. We had some fun asking about things that are coming in the future and one of the great people we're talking to, talking about resource in careers, said that in two or three years, we'd be with the farming unicorns, not hunting unicorns. I sat back and said, hold on, depends what unicorns and just have fun with this. And the point made with interesting language was that if you think three years ahead to your point on the kind of people you need is even more likely that you're going to find them in the market or that you're going to be guiding them and creating a career for them starting now, that ends up with the NatWest Group or anybody else, but you in particular, ends with you actually having a great career track that creates the kind of personalization that you need more of this. And your example on the fuller stack, but I promise you now, that if you're working in the healthcare industry or in the researching molecules, you don't want any risks. That's the thing you don't want. You want the specialists. Or your working gas and oil and you're trying to figure things out. And that was the point.

So what do you think about the careers. How do you create careers to serve you or do you still buy from the market? How do you see that going?

ZACHERY ANDERSON: We're definitely buying from the market right now 'cause it's the skill set that we're adding to the mix, so we're buying from the market. But I agree with you that in the long run, we need to be constructing careers.

And luckily, one of the core elements that we have of our strategy at NatWest and our purpose even is learning and education. And so, we have an amazing data academy at NatWest, which we're tweaking and evolving all the time, but one of the tracks in that learning academy is data science. And then, we also have career paths specifically designed for data science. And those include ML in them and I think that's really important. So I think we need to be and we're constantly hiring from university and pulling master students and bachelor students and PhDs.

And what I want is them to be able to have a career, maybe not like step-by-step every step career, but at least a learning journey that they're on personally, that the bank is facilitating and helping them with. And I think it's just really important. It's by the way, I think, also when I get together with a whole lot of other leadership around the data and analytics space, everybody complains either the rising cost of data scientists or the scarcity of inputs or how fast they move and all the non-techie places will complain about Facebook and Google, paying and stealing the best people and the techie places we complain about each other stealing each other's people.

But the reality is just that, look, data scientists – to be a data scientist, especially one with a graduate degree, you know, what their core mission is in life usually? Learn. It's kind of dumb to go to graduate school. And I mean I say that jokingly of somebody who's done it. It's really not a great economic decision usually. But you do it because you like to learn. And so, as an employer, if I want to retain my people, I have to give them a chance to learn. I have to be the best place where they can learn. And so, we do that through formal education and through create environments and allowing them to do that. I think it's just actually – it's one of those great places where purpose and good business aligns and it's good for your people too.

FERNANDO LUCINI: Yeah, I'm pretty impressed because I'm telling you, not everybody is quite there yet. And I will tell you, I keep on saying that data science – by the way, when I say data scientist, I mean I'm machine learning engineering, for me, they're the same space. It's a profession, not just a skill. So we got to respect it as a profession, otherwise, like everything, I can't wake up in the morning and decide that I'm a surgeon and I'm going to go and treat people or I'm an architect or I'm going to make a house. To some degree, stuff like this that requires advanced mathematical knowledge and that's not to say that you can't finish with a wonderful history degree and over the years become one. I'm not saying that. But on average, it is a profession that requires a particular background and it requires a dedication. And one of the characteristics that I think you said very well, is that learning, because the field is so broad that this continues to be evolving. I don't know if you've been tracking it, but in the world of natural language processing with the wonders that people like GPT3 and all these amazing people doing all these amazing things and there's not enough time in the day to read the papers that everybody's writing. All that changes, which actually, I'll tell you what. Here's a question that I think you and I have asked ourselves and worth sharing with the rest of the world, which is the idea of the role of the senior data scientist. Because I think you and I have very different views on this.

Because, of course, for me and I'll give my view and I'll let you give yours. For me, a senior data scientist, certainly in Accenture and our customers on my observation, is one that does have deep expertise on – or has had deep expertise in the field of data science. I'm sorry, it's so wide, that you can't be perfect in all this – has a deep expertise in one and has developed over the years a generalization of understanding, a broad-based understanding of all these such that he or she can apply very clear from business problem to algorithm to data can guide the rest of the community to what the good answer is. And I'll do a little addendum and I'll let you report back.

Which is funadementally enough, in a world of auto ML and all this stuff is going to make the dangerous amateurs I call them easier to actually press buttons and get an algorithm and see what happens. It feels that in a world a couple of years ahead, where you probably have many interconnected models to solve your problems. So they're all interconnected in a way that you pull a lever on one and three of them get affected or 100 or 1,000. The senior data scientists will make their value, if you want to put it that way, by understanding that ecosystem of models and the ability to balance these things in a way that, yes, you pull a lever on here and it has an effect on all others. And that that will become a skill. The balancing of all these models and, again, you have feedback loop on one that affects the feedback loop on another thing, you put another and so on.

How does it feel to you? What do you think the role of that lead data scientists or seeing a data scientist in that place?

ZACHERY ANDERSON: Yeah, or principal data scientist so that they can be used at EA. I think there's a couple roles for it. So I think you've described half the role. So I would expect a principal data scientist to be contributing the community of practice in your company. That means setting standards. That might mean we actually did some signoff work with our principal data scientists where he would take a look at – he might not be involved in the project, but at the end, he would come in and do some inspection of code, of outputs, question the team and kind of be almost like a certifying authority in some cases. And so, that's powerful.

And setting standards and tools and often, I would say that's the other place where I used my principal data scientist was in making sure that we were thinking deeply about what was coming next. The teams are using the tools that we provisioned, but maybe the principal data scientist might be pushing a particular direction somewhere or working with a small group, so we're pushing a particular direction where we want it to be a little more on the cutting-edge of something.

And then, I think the other role though is that a principal data scientist, because they have that capability to talk to business stakeholders, help people understand the models that are being used and see down to code because of the experience that they've had. I think they can also play roles leading or driving really big projects and they might be the one that is – you know, I think that role should have a large – should be able to say, alright, this set of models is critical for the outcome of the business. They're being developed right now. I think I'd like you to go run that and be the principal person on it and make sure that it happens, not from a delivery point of view, from an early on schedule or scrum master point of view or something like that, but from a like, hey, this is really important, this is something that's going to move the business if we get it right. I need you to get it right.

And so, I definitely use mine for special projects like that. Like, hey, we're working on some – I mean at one point, we dropped him into some work that we were doing on the AI for battlefield and that was a really powerful project. I mean he still did his kind of governance stuff and I mean he's over here with it and we said, like, hey, next battlefield's got to have great AI and we want somebody to jump in and dive into that project.

I think it's both of those, but what I'm surprised at often is companies don't have – and when I teach sometimes, I'll talk to people about this, but a lot of companies don't have an individual contributor, data scientist role. They don't have a senior role that goes all the way up. You either you stop somewhere and you become a manager and you can't advance anymore if you're a senior and I think that's a shame. That's not the right thing to do, either for the career or the person. They should be able to advance based on their impacts, not just on the number of people who they manage.

FERNANDO LUCINI: It's interesting because I see in contrast quite a lot of individual contributors who do not seem to be trying to help the firm going in that particular direction. So I think it tends to be a balance. And one of the questions I had for one of my guys who was lead scientist and I said to him, sometimes you got to love how academics simplify things. But he was an academic for many years. And I said to him, look, if in three years what's your job? And he said, that famous phrase, you know, if you think about it, it's a car analogy. It's not like we will be building the first electric car. At that point, we will be building the better electric car. So I said, what do you mean by that? He said, well, if you think about how a bank works and, again, let's not particularly in present company, but if you think about the kind of modeling he used to deal with customer loyalty using and I know those are point models, but particular types of models to solve particular problems, and you think about things like the GPT3 and all those other things, but now can understand and structure information in a way that's substantially better than before. His view was, my world will be to some degree to get the bang, the oil company, the whatever, to move away from the old and moving to the new, and I asked him, look, that sounds like usual technology power. Well, let's just knock the old stuff out because the new stuff is better and you don't know – always know what is better and the truth is science. Data science is not quite there.

These changes in science do sometimes bring very substantial benefits, yet, again, like any technology program, you still need to invest in changing the damn thing and turning it around. And I'm not quite sure I agree with him on this. I'll be honest with you. As I say, you assume the argument, I felt that that's a traditional IT engineering problem we've always had. And it felt like to me like the principal data scientists can be much more influential than just change management and change control and updating things and creating new.

ZACHERY ANDERSON: I think it would be odd if we had a lead data scientist as change control. That seems like a non-winning position.

FERNANDO LUCINI: And I say it that way.

ZACHERY ANDERSON: Yeah, but it's okay, I understand what you're saying. I mean I don't know. I mean you've been around as long as I have. These things have evolved. There's always new on the cutting edge. There's always new things to work on and incorporate that haven't yet done build into your tools that 90% of your population's using. And, yes, it's a technology point of view, but data science is a technology business. Our tools are evolving every – I'd say, we used to say five years, now I'd say three and it's accelerating. So the reality is the tools are evolving the modeling techniques are evolving. Those modeling techniques are allowing creative solutions to things that we didn't even think were in the set of things that might model in the past.

And so, even if you got a principal data scientist that can talk to the business and understand the business and then think about what we can do, if they know and are working on the latest techniques, they're going to be able to see opportunities that nobody yet knows are possible.

FERNANDO LUCINI: And the leadership are more educated as well. That's another of the bets that I make which is give it a couple of years, two, three years, and their average knowledge from folks of what data science can do and can do with the mathematical principles of statistics, just to make it really simple, will become substantially higher.

So rather than having leaders say, look, we're going to do AI this year. They're going to have a much, much better understanding of what kind of problems they want to solve, so the principal data scientist as they come through, they're validating an already kind of a little bit less shaky ground, stronger ground and suddenly, the value becomes easier to get around.

ZACHERY ANDERSON: I mean it's amazing. It wasn't so long ago, that we were all working like mad to build really good churn models for companies. And now –

FERNANDO LUCINI: We're still doing it.

ZACHERY ANDERSON: And we're still doing it. Of course, people are building churn models. But now, we're – that's what kids have coming out of school and that's what they teach at B schools is people to code and write churn models in Python, right.

FERNANDO LUCINI: Yeah, yeah or NeuroNets, right?

ZACHERY ANDERSON: NeuroNets, Yeah, NeuroNets. (inaudible). And so, let's not fool ourselves and at the same time, we've got things like GPT3 happening. So there's always stuff moving at both ends. The tools and the education is trying to keep up with the movement and then, there's people pushing new things. And I think that's, frankly, if that we're true, I wouldn't be in this. That's the fun of it right there.

FERNANDO LUCINI: As the last piece of the conversation, let me throw out a question that I ask myself all the time which is this, if you were to choose one thing that's going to be substantially different in three years, you can choose two, but if you had to choose one in our field that's going to make a massive difference to the field? It can be anything technology, how we deal with bias, you can choose it. What do you think is that, the one thing that you think matters to the future of data science, ML, engineering, etc., what would you pick?

ZACHERY ANDERSON: I think there's maybe I would pick two. It runs on the theme that we're talking about – we've been talking about here. But right now, the general population is increasingly uncomfortable with models that they can't understand, that they can't logically get their mind around. And as a result of that, we're doing a lot of stuff on AI explainability and I think that's really good stuff and we're developing – I mean in the bank, we're developing some great techniques around that in conjunction with (inaudible).

But I think people are just going to become more comfortable with the idea that the algorithms finding relationships and things that aren't understandable by humans and its performant and so, we can accept that it will do that – or that we will accept to be able to put it in the process. People will become more and more comfortable with that idea. Not in all processes, but in more and more.

And I think that's the game changer because that unlocks in some ways, we are the limitations of the application of these tools because of our lack of comfort around them in various places. And then, I do think we'll also develop more and more techniques and capabilities to really evolve fairness and embed it in the model some more of the human values that we want to have that we're afraid of losing in the models and that's a move evolving really quickly.

The other thing that I'd say I think on the ML front, and we both know this, but bias in training data is the problem. The models are working really well, increasingly well in lots of places. But bias in training data is the issue and unknown bias. So that's where I actually think this area of fairness will go is using these same techniques to understand the bias as opposed to just find relationships.

FERNANDO LUCINI: Yeah, I like that one. I'll tell you one that I also throw in for you for good fun, which is that if you think the amount of models that we have that fail having the right amount of data to do what we want them to do, large amount of models end up having that problem. I think in three years, these new techniques around synthetic data iteration is a good one, you know, the DAEs and all that kind of stuff. The iteration on term coders is a lot. I think there's going to be an entire economy around, let's say, NatWest – and I'm making this up – by NatWest being able to create a synthetic data set, none of the patterns, none of the actual - or none of the actual data itself, but all the patterns piling together with other banks, for example, for to find crime. So I think there's going to be – there's a potential and I hope it happens, a potential of having an entire real push back to the POCs if we can use synthetic data for all our POCs, most of our stuff will go to production because that's one of the problems. It takes us so long to get the data, we end of getting the wrong data in the first place.

ZACHERY ANDERSON: Yeah, but we never get the data we really want because of a privacy concern or something like that.

FERNANDO LUCINI: Exactly. So I have to think that maybe in three years, synthetic data will be a big thing which will bring all sorts of other fun economic issues, such as would somebody spike the data to – we digress. I quite like that one.

ZACHERY ANDERSON: I think that's a good one. I'll buy into that. I think you're right. I think we are still early days on synthetic data and I think there's a lot of development there that's plausible and a lot of interesting places looking at it too.

FERNANDO LUCINI: Oh, yeah, and it's outside the big companies are using it for reinforcement, learning and stuff, which is the more traditional thing to see. Let me give you the other one that I quite enjoy as a lateral thinking which is if in three years, we're going to have the ability to pull a model, make it fit and do all these things pretty easily, almost as I say, at an amateur level. It's going to move much faster than the governance of that process. There's going to be a real place where we can create – we can bring data, we can bring model, we can make the thing fit to an objective, but when it goes wrong, not as much of a greater understanding as to why it went wrong.

ZACHERY ANDERSON: Yeah, well, I mean I don't think we have to wait three years for that, I think that exists now. But, wait, you don't have to be doing ML to make that mistake. You can use regular old stochastic models and get that wrong at various levels of complexity.

FERNANDO LUCINI: Correct. So I think I'm banking on the multiplicity. So we're so wanting to get all those algorithms because everybody wants to play, then we're going to be in a place where following behind are going to be some of us wanting to say, hold on, you still need to understand how the NeuroNet works. Because if it doesn't work, you want to know why. And I think that's going to have an interesting side effect. And that's why I don't know exactly where it's going to go. Nobody knows. But I still have a great feel, but it feels that's the reality. The one is going to go faster than the other because we all want to run. We all want to run, but then, and there's going to be a place for those performance data scientists who come in and say, let me tell you how all this thing hangs together.

ZACHERY ANDERSON: It works together. Yeah, I mean, again, those are the same problems that we see even in stochastic models, right. Stochastic models stacked on top of each other with correlated errors. You don't have to look very far to find that.

So I think you're right. I think that problem exists. I look at some of the competitive models being built like these nested reinforcement learning models where there's two models competing to fool each other and you think about the implications of things like that for governance. With creative thinking, we can build these tools to governor also, not just to solve a business problem and I think that kind of broad thinking is starting to happen more and more too. So it might be a little behind, but it's actually increasingly – I think it's an increasingly interesting space.

FERNANDO LUCINI: I'm with you. And on that bombshell, 'cause you and I could be talking about it.

ZACHERY ANDERSON: We can talk about this for a long time.

FERNANDO LUCINI: For hours. So thank you very much for coming and hope you actually had some fun running around this topic, which as I say, we could be on it for hours and hours and hours. But thank you for the time and no doubt we'll get some commentary and you and I will have to answer with some of our examples.

ZACHERY ANDERSON: That's good. I hope I haven't committed any crimes or anything.

FERNANDO LUCINI: Crimes against data science and engineering, right.

ZACHERY ANDERSON: Crimes against data science or in my case, the more likely one is I'm going to commit a crime against software engineering.

FERNANDO LUCINI: There you go.

ZACHERY ANDERSON: Hopefully, you've kept me in line there.

FERNANDO LUCINI: Thank you very much, Zach.

ZACHERY ANDERSON: Same to you.

FERNANDO LUCINI: See you later, bye.