Me, Myself, and AI Podcast

Trust and Fraud Detection at Scale: Instagram's Stephanie Moyerman

SAM RANSBOTHAM:
- The emerging field of data science is progressing rapidly, resulting in numerous managerial challenges. How do technology leaders ensure quickly evolving digital spaces like social media stay safe and equitable? Find out on today's episode.

STEPHANIE MOYERMAN:
- I'm Stephanie Moyerman from Instagram, and you're listening to Me, Myself, and AI.

SAM RANSBOTHAM:
- Welcome to *Me, Myself, and AI*, a podcast on artificial intelligence in business. Each episode, we introduce you to someone innovating with AI. I'm Sam Ransbotham, professor of analytics at Boston College. I'm also the AI and business strategy guest editor at *MIT Sloan Management Review*.

SHERVIN KHODABANDEH:
- And I'm Shervin Khodabandeh, senior partner with BCG and one of the leaders of our AI business. Together, *MIT SMR*and BCG have been researching and publishing on AI since 2017, interviewing hundreds of practitioners and surveying thousands of companies on what it takes to build and to deploy and scale AI capabilities and really transform the way organizations operate.

SAM RANSBOTHAM:
- Stephanie, thanks for joining us. Tell us a little bit about your job. What do you do right now?

STEPHANIE MOYERMAN:
- I am the data science director supporting Instagram Wellbeing. Instagram Wellbeing ensures trust, safety, and integrity on Instagram's platform.

SAM RANSBOTHAM:
- We've heard a lot of guests [talk] about the various dark sides of social media. It seems like there's a story almost every day about something that's gone wrong. I think we overlook all the things that are going right. What kind of role do artificial intelligence and data have in helping to bring the light back to social media?

STEPHANIE MOYERMAN:
- There's a wealth of data that actually show that across the sweeping industry — and not just social media but across a general industry for teens — safety over the last 100 years, well-being over the last 50 years, has actually increased for teens. And there's an interesting data story on that, if you look at everything from what you think of as your most severe potential problems, such as police reporting, to just the ability to access information to teach yourself ... and really as simple as literacy rates have just increased time over time again, and it's really the data that gives us the lens on that. So being able to analyze the data on a platform like Instagram or on any of these broad social media platforms, where you have access to tens of millions of teens, and say, "Hey, this is how we can actually tell a story that makes this platform safer, better, and makes the world a slightly better place," is something that, previous to the last 20 years, we actually just could not do at all.

SAM RANSBOTHAM:
- Hey, that's a great point. I mean, we just didn't have that data at all in any way. There was no collection of it at all.

Tell us a little bit about what you learned about how things are progressing. You said literacy rates; what about other things?

STEPHANIE MOYERMAN:
- Within our own ecosystems, where we see things progressing is in the ability to help set healthy social norms, and this is how we think about things. Not just for teens, but for the platform all up, how do we encourage and foster relationships among people, and of course a dialogue that helps establish a normalized behavior that increases awareness of the good that's going on and sets a tenor for community conversation? And all of this is driven by large-scale personalization models — models that try to predict what it is that will elicit the appropriate responses to a given situation.

On the opposite side, the entire industry, actually, across e-commerce, across social media, has large-scale models of AI that have enabled us to understand the sentiment of conversation, the safety of conversation, the content within images, that protect people from seeing things that otherwise would have gone unregulated. So if you think about sort of tabloid publications or things anyone could throw at your door before that were largely unregulated, now we actually have not only the breadth to do this, but the sweeping AI tools that allow us to do it at scale.

SAM RANSBOTHAM:
- That's beautiful, because you're pointing out how so much of this stuff just was happening beforehand, and now we recognize that some of it's just a measurement and an observation problem.

SHERVIN KHODABANDEH:
- That's true. It also feels to me that there are two factors going on, and maybe there's a race here, because there is, on one hand, more and more people [with] access, and this is a sign of a free and thriving society, and people should have freedoms to post things that they want or share things they want.

On the other hand, the chances of bad or hurtful content has increased because there are just more participants. So that's one factor going on.

On the other hand, I think the point you're making, Stephanie, is like, "OK, so now we have data, now we have AI, now we have ML [machine learning], so we can catch it." But it also feels like it's a little bit of a race. It's like in health care, too, our life expectancy is increasing because we are living longer because of medicine intervening. But then also you hear about all kinds of sort of new pollutants that we're inventing, because we're evolving and we're inventing these things. But because we are also evolving with medicine, we're trying to diagnose and catch that.

Does that make sense? It feels to me like there are two different factors, and there's a race here, or is the race going to end at some point and AI and technology are going to rein [in] the bad elements?

STEPHANIE MOYERMAN:
- You know, I think it's an interesting question. One of the big things that we talk about frequently is that all of these spaces — spaces [like] integrity, medicine — where you're trying to battle this unforeseen negative is that they're are

adversarial games; they're cat-and-mouse games. And so the more that you shift, the more creative are the differences that you create in the ecosystem of the people you're battling against.

But in order to keep up with this, actually, on both sides, the technology has to get more sophisticated, and you have to be able to scale more quickly, and you have to work through a different set of issues than even we would five, 10, 15 years ago. And so this is sort of like a bit of an arms race in technology.

What I would say, though, is at large scale, at massive scale, there's the ability to regulate, which is extremely important for setting a universal standard.

That said, I actually love the take on freedom and agency of expression that this gives. I think, having worked in this field for a long time, (1) the adversarial nature is one of the things that keeps me in it — there are always new problems to solve, there are always new things to work through; but (2) I actually do believe, even working in fraud, abuse, [and] trust for the past 10 years, I believe that people are fundamentally good. And we say this a lot: A bad actor does not mean you're a bad person. Someone can [demonstrate] bad behavior without realizing it, without the intent that you think is associated [with] that.

So I think these large-scale efforts around understanding tenor of conversation, understanding the motivations for some of these efforts, and pairing things like AI with social psychology, with government regulation, to allow people to express freely their chosen agency of experience but just make sure we have the proper safeguards, and at scale, to prevent anything that's massive from happening incorrectly in the opposite direction.

SHERVIN KHODABANDEH:
- Yeah. And as you've been saying, we now have amazing training data sets so that you could do that. I always wonder, though, with fraud also, payment companies eventually figured out that there's a lot of data they could [use to] sort of understand fraudulent behavior, and they trained the data, and if the models were wrong, they learned because there was a source of truth, and so this was fraudulent, and it's a fact. How certain can we be about the source of truth on things that you're talking about? So I mean, how definitive is the data to train these algorithms?

STEPHANIE MOYERMAN:
- You know, this is such a good question. I think the truth is, the models that we teach with our training data can only discern as well as any human could discern that's an expert in that field, right? So the models can look at many, many, many more examples of this — millions more examples than a human could in a lifetime. But the humans have built their acute knowledge of what constitutes, let's say, fraud, what constitutes good art, and we've pumped this in. The model is learning from those decisions.

And so there's two really important things here. One, I remember working on a project once where somebody was saying, "I want to predict sentiment through viewing you through these glasses." And a guy on the team said, "If you can do that, can you give them to me? Because I still cannot tell what my wife is thinking." And so, you know, the ability for us to do that is kind of gated by foundational human knowledge.

I think a lot of it is cultural, too. If you think about what constitutes a product or what constitutes an image that somebody might find risqué will vary wildly around the globe. And so you have to put all of these models in context. But the most important part is the one that you touched on earlier, which is, there needs to be a closed feedback loop, and that's how fraud really works and why it's so good: because there's an agent checking when there's money on the other end of the line to ensure that this actually was fraud, checking with the person who, let's say, owned the original credit card that was, in this case, in this example, stolen, right? And so you have this closed feedback loop of all players.

That's not 100% guaranteed either. I mean, if you mess that up, that's insurance fraud, but there are these closed loops that say, "Hey, our algorithm got this one wrong. We need to fix it." And so for all problems, if you want to evaluate them in context, you need to be able to have a real-world feedback loop and ground truth that's continually being iterated on and has that agency and voice for reporting. So if you take that away, now, all of a sudden, the algorithm is only learning from itself. And that will be a problem.

SAM RANSBOTHAM:
- That's really interesting too, because I think both you and Shervin mentioned appropriate or bad behavior, and these are such subjective types of labels that can vary a lot. And I think what's interesting, you know, as we think about how this data science field is developing, there was a quote a few years ago that "the sexiest job of the 21st century is going to be the data scientist." Well, the kinds of things that both of you are talking about speak to the idea that the next sexiest job is going to be philosopher. It's going to be that person who can reason out what these algorithms should or should not be doing, and that feels really scarce to me right now. Data science ...

SHERVIN KHODABANDEH:
- Isn't that what Plato said 2,500 years ago?

SAM RANSBOTHAM:
- Cycling back to the beginning here? But that does seem like a skill that's much scarcer now than data science — this idea of determining, like you just said, whether something is appropriate, what those norms should be. That strikes me as really hard.

STEPHANIE MOYERMAN:
- It's a really interesting thought, because data science really didn't exist 20 years ago, right? So the number of people who can practice data science at a sophisticated level, it's very small, because you had to grow up in a world where you could learn it and have 10 years of real-world experience on your resume. "We need somebody with at least 10 years managing data science." Good luck; there's like seven people out there, right? Like, who are those people?

And so when you think about that, the notion that these algorithms have gotten so sophisticated that we need a lens on fairness, on understanding intent and not just detection of behaviors, on interpreting what is an appropriate action, that AI can drive versus just what's a decision, a judgment in passing that it can make — that's really new; that's like last-five-years kind of stuff.

Now we have very sophisticated user experience researchers that can understand the philosophy of these products, that can help to move things

forward. I think we need to, exactly as you're saying, encourage people not just for the data science or the technology aspect of this, but to go into the field of learning what it means to be in an AI-driven world.

I think where this affects … most importantly where we do have the most experience right now is actually in fairness — so in saying, "Are AI algorithms fair? How do we slice the output of these algorithms across different demographic groups?" Demographics could be any way in which you want to slice it, like Shervin was talking about earlier — like culturally, to make sure that we have equal outcomes for all parties, or that we adjust algorithms so that we do automatically. And there's a suite of software out there that is excellent at doing this, even at just the start of this burgeoning field. And so we need experts to sit with that suite of algorithms to help companies figure out how to traverse these super-sensitive and very, very important topics.

SAM RANSBOTHAM:
- Like you said, fairness may be one that we're more developed on than other aspects, but that's challenging, because right now people are posting pictures on Instagram and other places — you don't have those five years or those 10 years to develop that experience. How do we do that in real time while the horses are escaping the barn?

STEPHANIE MOYERMAN:
- I think that the issue here for all of these major platforms [is] when you see this at scale and you say, "It's both exhilarating and it's a little bit frightening." And I understand the consumer aspect of that, but we would never have been able to get to this scale already, where Amazon can deliver you a package in a day, had we not had all of these algorithms vetting everything they're selling, keeping them compliant, vetting every single purchase to make sure no one's stolen your credit card, vetting every single thing to make sure the platform stays safe and healthy.

But really, the issue for me is that, as we talked about earlier, you have to close the doors on things to ensure there's safety and protection on the platform. You have to develop these large-scale algorithms, and you have to have a set of human judgment labels that tell you what's appropriate versus not. And then you have to have your systems designed so that your consumers can give you that constant feedback and very easily — again, the agency. We have to enable our customers, our users, everyone to have a voice and tell us, "Was this a proper decision or not?" and support that through operations on our side to say, "We have the cleanest set of decisions that reflect the tenor of the community that say this is acceptable or this is not acceptable by our standards."

SAM RANSBOTHAM:
- Actually, what I liked about that is the analogy you made with payment fraud in that … I don't want to imply that problem is solved, but, again, as we got really good at that, it's easy to get it to scale. You can have that state-of-the-art detection really scaled throughout the organization quickly, and what you're pointing out is that as we do this with other decisions and other aspects of the platforms, the great thing is that we can push all those out at scale as well, so when we do make an advance, that advance shows up, really, everywhere — quickly. That's big.

STEPHANIE MOYERMAN:
- I think one thing, too, is that there is this giant fear of, sort of, these large platforms in AI that's kind of circulated, and I

understand why: The scale [of them] is almost incomprehensible, right?

But I think one of the things that would help is if the community generally had a better understanding of how their actions actually affected the way in which we thought about these algorithms running. So I had a chat with someone a long time ago, and they were arguing with me, and they said, like, you know, "I don't trust AI decisions. I don't trust these machine learning decisions. I'd rather have a human in the loop." And I said, "I actually trust AI decisions more than I trust human decisions." And they said, "That's because you understand how AI works." And I was like, "That is why."

I would argue, being a nerd, that I probably understand how the algorithm works better than I understand how the human brain works for most people. And so I think if others leaned in a bit more to understand how their actions on all of these platforms — clicking "I want to report something," clicking "This is an incorrect decision; I'd like to appeal" — how much that influences. And the ability that they have to provide accurate, adequate data, the better actually they can proactively make our decisions in this large-scale world. I think people ... it's just like voting, right? People will say, "What's one vote? Why does it matter?" But when everyone gets together, you actually do have a say in shaping the outcomes for these large decisions.

SHERVIN KHODABANDEH:
- I love the voting analog: that we get together, we vote, and we make change happen. So we usually vote anonymously, and we fundamentally trust the system, otherwise we wouldn't vote. But it feels like you need an external force. I don't know whether that resonates.

STEPHANIE MOYERMAN:
- I agree with this completely. I think, to be honest, if you look at what Apple's done with the most recent updates, where they ask you, "Do you want to track?" and they're very overt about it, I think this is a step that generally regresses the availability of data in the short term but actually increases the quality of the data in the long term, because volunteer data is data that you know is good data, generally, that you don't want people to have to infer.

So it gets you something. And having incentives aligned with having the best algorithms is something that helps. It's the same way with the voting analogy. If you feel like you are actually incentivized to use your voice and cast your vote on the ballot, you generally will go to greater measure to make sure that you're informed about what it is that you're voting for.

SAM RANSBOTHAM:
- Let's go back to your understanding of the algorithms. How did you get into this role? Actually, my father-in-law's a nuclear physicist, and so I'm really hoping you can say something cool and physics-related right now so that he doesn't think I'm a crazy man like he already does.

STEPHANIE MOYERMAN:
- So it's probably inferred from the question: My background is actually in cosmology and astrophysics. I have a lot of friends in the same field who have all gone into data science and machine learning.

If you work in data with physics or engineering, sensor data is huge. It's a massive data set streaming, so from our telescopes that we set up in Chile, you're pulling 500 hertz, 1000 hertz samples per second off of these sensors, and you've got thousands of sensors and you're running

them 24-7 and trying to process this into something coherent.

And if you think about the way social media or e-commerce works, it's the same thing. It's just signals flowing in from all over all the time, and you're trying to process this into a set of coherent decisions. One of the biggest things, though, is almost any scientific endeavor that's large-scale now — nuclear physics is one for sure, particle physics — it's about finding the signal in the noise. So you get so much noise. The cosmic microwave background that I studied in graduate school … the signal we're looking for is one part in 10 billion to the noise.

And if you think about things like detecting fraud, detecting images, you're not talking about 99% of these coming through are the ones you want to pull out. It's quite the opposite. And so really, your good signal is almost like noise for detecting these anomalies. And that's a lot of what data science with regard to physics is: You set up these large scale systems to find these very, very tiny signals that indicate something about the origin of the universe or how particle nature is formed.

SHERVIN KHODABANDEH:
- That was very poetic. I think that would make your father-in-law quite happy, Sam. Direct him to that section. That was very well said.

SAM RANSBOTHAM:
- Finally, finally. Grandkids later, this may be the thing to push us over the edge. But I do like that that thinking and, you know, may explain why we see people from these disciplines, like physics, who are used to processing these large streaming data and picking out that tiny signal in there as being a valuable skill here in modern commerce.

Kind of looking back on my own reverie, I actually got started and interested in this looking at security logs, where there's just billions of records and only a handful are bad, but figuring out which ones are bad is how I actually got started in learning some of these skills and some of these tools. I think that's a really fascinating analogy. Shervin and I are both reformed engineers too, so that appeals to us.

STEPHANIE MOYERMAN:
- Just an anecdote about this. We were talking earlier about the newness of this field, right, and how this just wasn't available so long ago, and when we were in graduate school building the telescope, we had an internet link — a very fast one — from the telescope site in the middle of the Atacama Desert in Chile, but it wasn't fast enough to actually get all the data to Lawrence Berkeley's supercomputers to process, except for having it way down-sampled.

So we down-sampled like crazy just to make sure everything was going well, produced these intermediaries, and we had to invent a new file transfer protocol that we called HDOA, which stood for hard drives on airplanes.

We would literally be flying back with suitcases of hard drives from the desert to upload that data so that everybody could use it. And, like, if you think about how far

we've come, just, you know, 20 years later, you can see why people from these fields that were doing hard-drive-on-airplane transfers are now the ones that are working in this heavy data science realm.

SAM RANSBOTHAM:
- All right, Stephanie, now's the time we have a series of rapid-fire questions, so just answer the first thing that comes to your mind.

STEPHANIE MOYERMAN:
- Oh, God.

SAM RANSBOTHAM:
- What are you proudest about that you've accomplished with artificial intelligence?

STEPHANIE MOYERMAN:
- We actually did a livestream — this is not related to fraud, abuse, [or] trust at all — we did a livestream integration with the X Games many years ago. We actually put tiny sensors on the [snowboard] and classified the tricks and the hang time of the athletes in the winter X Games in real time. That was the coolest thing I've ever seen happening onscreen in front of me and in real life at the same time that I cannot express how much awe I had in those moments.

SAM RANSBOTHAM:
- That's cool, because whenever I see those commentators, they something like, "Oh yeah. That's a quadruple whatever," and to me it was just a giant blur.

STEPHANIE MOYERMAN:
- We actually had to change to make sure that we were not trying to guess spins and flips, but rather rotation around this axis and rotation around this axis.

SAM RANSBOTHAM:
- All right. So we mentioned bias and ethical issues, but what worries you about artificial intelligence?

STEPHANIE MOYERMAN:
- I think the lack of very experienced practitioners, actually, is my biggest concern in this space right now. So if you're thinking about it, if your children are thinking about, it just push them further into this field.

SAM RANSBOTHAM:
- Good. What's your favorite activity that does not involve technology, that's not AI?

STEPHANIE MOYERMAN:
- I am a very experienced judo and jujitsu player. I have been doing judo since I was four. My dad's managed two Olympic teams for judo, and so it eats a lot of my time outside of hands-on keyboard time.

SAM RANSBOTHAM:
- Crazy! All right, so what's the first career you wanted to be when you grew up? Judo master?

STEPHANIE MOYERMAN:
- When I was five, I wanted to be president. I'm so glad that one didn't stick.

SAM RANSBOTHAM:
- Yeah. That's not a ... that's a tough job. What's your greatest wish for AI in the future? What are you hoping we can gain from this?

STEPHANIE MOYERMAN:
- I would really like to see AI applied to some of the world's most systemic problems in fields that are a little bit slower and more nascent. So I think things like the work that the [Bill & Melinda] Gates Foundation have done in trying to create global equity and solve food-

shortage problems or disease problems. I would love to see a lot more AI lean in these areas, particularly when it comes to things like hardware distribution channels, so that we can actually effect global change in areas where historically we've been unable to scale.

SAM RANSBOTHAM:
- Actually, that resonates with your overall theme of scaling. Stephanie, I think the things you're mentioning about how the ideas of us learning at scale, both getting the lessons that we learned from artificial intelligence throughout organizations and also learning what the algorithms are telling us how to improve — I think these are some fascinating things, and we appreciate you taking the time to join us today. Thank you.

STEPHANIE MOYERMAN:
- Thank you so much for having me.

SAM RANSBOTHAM:
- Thanks for tuning in today. Next time, Shervin and I talk with Shelia Anderson, chief information officer at Aflac.

ALLISON RYDER:
- Thanks for listening to *Me, Myself, and AI*. We believe, like you, that the conversation about AI implementation doesn't start and stop with this podcast. That's why we've created a group on LinkedIn specifically for listeners like you. It's called AI for Leaders, and if you join us, you can chat with show creators and hosts, ask your own questions, share your insights, and gain access to valuable resources about AI implementation from *MIT SMR* and BCG. You can access it by visiting mitsmr.com/AIforLeaders. We'll put that link in the show notes, and we hope to see you there.