FTC FinTech Forum: Artificial Intelligence and Blockchain
March 9, 2017
Deirdre K. Mulligan
Transcript

DEIRDRE MULLIGAN: First, I just wanted to acknowledge the fact that the Federal Trade Commission over the past, I guess, 15 years maybe has played an incredibly important role in looking really deeply at shifts in technology, and how they affect the marketplace, and what are the consumer protection issues they generate. And we're rather unique in the way in which we've done this globally-- that they've provided this forum where we can come together, generate accurate information about how technology is changing, document the facts on the ground, so that we can have agreements and disagreements about the policy implications, or the risks, or the benefits, but that we have some shared understanding of what our facts are. And so in that framework, I think, it's incredibly important that you're looking at these technologies. So I just want to thank you for that.

I was asked to frame the issues, and we are going to have a rich conversation about benefits and risks. I'm trying to frame some of the consumer protection challenges, because I think it's helpful if we have some shared perspective on what those might be. And I'm going to talk a little bit about AI generally, but really want to pretty quickly drill down on some of the opportunities and risks presented here with AI in the financial services realm.

So first, I'm working on a project about handoffs. And what a handoff is-- we're literally taking our baton and we're handing off sensing, decision making, maybe acting in the world from human actors to technical actors. Or we're kind of reconstituting the socio-technical systems and redistributing responsibility for different functions. And today we're talking about this shift between this intimate conversation that you might have with someone across the desk to a form that you may be filling out in a chatbot that you may be interacting with.

And in those translation moments we had the first question we always have to ask-- are we really doing the same thing? It's really important to understand how the activities that we've done historically in one way are translated with some degree of functional fidelity. What does it mean for this thing to be replaced by this thing? And paying attention to that translation, the things that might fall in the cracks become really important.

So I want to talk a little bit about some of the values in systems. And you know like, well, this isn't about financial systems. But these are really great examples of the various ways in which bias can enter systems.

The first is a Volkswagen. We found out that there was an effort to intentionally undermine regulatory objectives by putting code in systems that allowed the cars to identify when test conditions were occurring and to go into a separate mode that altered the emissions profile of the car. And you might be surprised, but when David Wagner, who's a professor here in computer science, and I, and a whole team of other people looked at voting machines in the state of California, we found test-mode code on voting machines too. And so this effort to intentionally embed bad code to defeat regulatory oversight is one way in which bias can enter technical systems in a way that is very, very difficult for the public and regulators to understand.

The second bias can emerge through the process or values of designers. This is the Jeep Cherokee. You may remember that Charlie Valasek and his colleague were able to remotely take over this car, control the air conditioning, but, most importantly, the gas and the brake on a highway. And was that because Jeep doesn't care about security? No, it's that as computation and connectivity is getting embedded in devices, particularly big ones that move about, the engineers in those companies have not historically been focused on computer security, cybersecurity, and so it's a blind spot. And so we have to be concerned about the ways in which systems might just reflect the institutions or the individuals that are building them.

And finally, bias can arise through complexity. And I'm going to point to some more research that has been conducted here, at the International Computer Science Institute, where they were trying to understand why men and women might be seeing slightly different ads for job services. And it's really difficult to figure out why it's happening. Right, it could be because advertisers are actually intentionally selecting to target ads to or away from certain populations. It could be because the ad serving algorithms are watching people's behavior and responding to it in a way that targets ads towards women or away from women. Or it could be that people are just getting outbid by other actors in the system. And because these are multi-party systems, sometimes figuring out how bias might emerge, regardless of anybody's intent, takes a lot of digging.

So looking at financial systems, we see some similar issues. One, when we're thinking about the development of these algorithms and the technical systems, the front ends and the back ends, we want to be thinking about sources of intentional bias. And we know that there are incentive structures for financial professionals-- how people are rewarded in their jobs. There are also incentive structures for firms. Who are they affiliated with? What sorts of relationships do they have? And the selection of a portfolio of funds-- which things can you actually purchase with this particular tool? All of which you can think about as points that may produce certain sorts of biases in the design of the technical system.

Second, the designers' values can be a source of bias. I spent some time looking at a report that was done by FINRA, the Financial Industry Regulatory Agency-- I think I got that all correct. And it was super interesting looking at the front ends of the robo-advising technology in that the way in which they go about assessing risk are all really different. And they reflect a lot of designers working with financial firms, trying to figure out how do we do a good job assessing risk. But they're all very different. And they present another opportunity to think about bias in the system that they're going to be eliciting different sorts of information from people.

Second, the presentation of options. We know from behavioral economics, cognitive science that framing, ordering-- there's a whole host of issues around the representation of choices-- that can really influence the way in which consumers act.

Finally, the complexity-- we're talking about data, we're talking about algorithms, we're talking about interactions and personalization. These are multiple parties engaged in trading and buying.

And so there's lots of opportunities for bias to come into systems in super complicated ways that require things like game theory thinking.

Second, the data used in these financial advising systems can introduce sources of bias. So one, what data is considered relevant? As I was mentioning, in those intake forms the data that is requested and used about clients varies. There's some standard pieces that appear to be required, but there's a lot of variation.

What data is provided about the funds and what data is available to the investment advisers about those funds-- is it uniform? Are there gaps in it? Which brings me to the accuracy and quality. Are there these limited gaps in data? Is there some asymmetry between what investors may know versus what the firm or what other people they're bidding against might know about the funds.

The client input. Do clients understand where data is coming from? Some of it may be coming from them, some of it may be coming from third-party sources, which may be independent sources of inaccuracies or outdatedness. Do people understand the need to keep data updated? Are there incentives in place for people to refresh information?

How is data cleaned? This can be super important. People are often very focused on what data is there. But if you are a data scientist, one of the most important decisions you make after what data you use is how you clean it. And to give you a really poignant example, if you remember back in the 2000 election, a lot of the grief was caused by how the voter files in Florida were cleaned before they were used to purge people from the rolls. And if you weren't sensitive to the fact that the African-American population is far more likely to have people whose names are in the common set of names, and therefore are more likely to have false positives.

You are cleaning, removing things like Sr. and Jr., for example. Despite the fact that you had no intent to have a biased outcome, might have a quite biased outcome. Because you have to understand the distribution of the information across the population.

And finally, sampling bias. Who is poorly represented in the training data from which the models are built? Many of these services are being targeted at millennials, and yet those are people who have historically not been a big part of the financial system. And so by building models based on data that doesn't represent them, what are some of the risks that we might be introducing?

So thinking about some of the-- kind of an overarching level-- what are some of the values at risk as we move to AI? Generally, one is privacy. In order for the machines to learn we basically have become big sources of data. We're just emanating lots of information, it's all being collected. Super useful when we're thinking about personalization, and we're thinking about making our lives with machines more intuitive and less demanding about clicking boxes, et cetera-- great benefits. But there's a huge amount of data about individuals that's floating about.

Two, autonomy. People are experiencing lots more nudging that is subtle and less obvious to them. They are being advised, their information environment is being shaped. And some of that may be super positive-- we may be getting people to invest more. And some of it we may have concerns about-- undue influence, and steering, and other things that have historically been part of our consumer protection dialogue.

Second, we know from research that when a machine tells us something, people are far less likely to question, to ask about its priors, or its limits, or its background assumptions. And therefore it is taken as kind of ground truth. And this can be super problematic. As we know, it's really easy to lie with data. And just because it's mathematical or just because a machine produced it-- everyone in this room knows-- that doesn't make it true. We still want to interrogate all of those things. Unfortunately, we know in practice humans respond to machine sometimes in, well, that's just the way it is.

Fairness. The data that is sucked in is used to create profiles. And what characteristics are being used to ascribe those profiles may have some serious consequences, depending upon how they relate to underlying demographic traits in the population.

We also have to be really concerned about non-distributive group profiles. So the fact-- I may create a profile of people like you, but because those profiles are not equally distributed across the population, it actually isn't like you at all. At a statistical level it's just like you, but on the ground you look nothing like that profile.

And this can have an enormous impact on the profiles that are offered as defaults for people. And if your assumptions about what they are, and who they are, and what might resonate with them, what might be good for them are based on statistical profiles that are at odds with their individual needs, that can be super problematic.

It can also lead to this fact that people can get trapped by past behavior, and we've seen a lot of concerns around that in the criminal justice side. And it could be about group behavior, it could be about individual behavior. If we're profiling individuals in a way that narrows their opportunities, that's not necessarily best for them. It might be most profitable for us, but it might not be best for them.

And finally, responsibility. As we reallocate tasks between humans and machines, one of the things that can sometimes fall through the cracks is who's responsible for different parts of the function. And more importantly, perhaps, for regulators-- who's accountable. And we see a lot of different responses to this concern about accountability playing out in the automotive sphere. But it's really important here too.

So in thinking about these handoffs you want to be paying attention to what I like to think of as properties of the system. If we really want to think about functional fidelity, we want to transfer this function from humans to machines, what properties do those machines need to have to really make sure that we can safely and fully transfer that function or be really clear about what we're transferring and what we're keeping on the human side.

Two, people and algorithms as organizations. A lot of people think about algorithms as another form of bureaucracy. They're nice iron cage that we can build. And the question is, how do they shift power, and particularly in this context, power between financial professionals and technical professionals who build systems? And how do we make sure that the values that might adhere in one translate into the technology that is built by another?

How do we manage values and risks? We know that the financial sector can be incredibly volatile because of those three Vs-- velocity, volume, and variety. And that means it can fail spectacularly, as we're also really familiar with. And when we have lots of people who are now doing their own activity with less insight and forethought potentially by financial professionals, how does that play out?

And then finally, social input into the guts of the machine. There's been a lot of calls for transparency. We want to see the code. And I think most of the people know in the room that that's kind of a silly response. Looking at the code-- again, one of my colleagues here in computer science, David Wagner, was trying to see how well people who had security training did at finding bugs in code. It was not so good, none of them found all the bugs.

And so we do need methods to make sure that these machines, these algorithms reflect values that are broadly supported, not just the values of designers. But transparency isn't the answer to that question. So what is? Well, there are a few things that I think we can do right away. And we see-- the SEC guidance that came out last week I think was really useful. There's some really useful information in this FINRA report that I was looking at.

But one question is, do consumers understand the product and how it works? Interpretability rather than transparency. Do I understand it? What is the algorithm doing? How does it do it? What kind of algorithm is it? Who controls the algorithm? Because that might have a lot to tell us about those potential sources of bias. What data is being fed into it-- both the sources and the specific kinds? And biases-- does it favor certain funds? And if it does, why? And why might I care about that?

Finally, can consumers and financial professionals interact with the system safely? And what do I mean by that? I spent a fair amount of time looking at the disengagement reports and accident reports from self-driving cars, autonomous vehicles. And it's super interesting to watch the way in which what I like to call not really the handoff but like the throw back or the grab back-- like the car is going along, and then I think it's doing something wrong, and I grab control, and then often I get hit, if you look at the disengagement reports.

And this question about how we structure that handoff between humans and machines, so that we do no harm, is super important. And some of the things that it requires is exposing the assumptions and limitations of the models that are embedded in the algorithmic systems. And we saw this in my mom's response to the fact that a Tesla car drove into the side of a truck. It was like, I thought that thing could see. Like, it just rode into a truck, Dee. Like, what does that mean? And I'm like, mom, it doesn't see like you.

Seeing for a car means a totally different thing. And if you understand how it sees, you might understand why it was possible for it to literally drive the person under a truck. But if you don't, you think it's OK to let go of the wheel and watch a Disney movie.

So this handoff between algorithms and humans, and making transparent where out-of-bounds actions might be taking over-- because we see some serious fluctuation in the market that we're not sure our models are going to be able to ingest-- are super important. And to get at that kind of interpretability and understanding we need to think more fully than just about disclosures.

We need to think about simulations. The way in which people learn about systems is they play with them. Like, what happens when I do this? And so I think simulations, and modeling, and other ways that consumers and financial professionals can play with these systems-- they can get a deeper understanding of how they work and why-- are super important. And, of course, making sure people have access to their data and understanding of how they're being profiled. So with that I'm going to turn it over. And this work is funded by the National Science Foundation.

Thank you. [APPLAUSE]