Stephen Piepgrass:

Welcome to another episode of *Regulatory Oversight*, a podcast that focuses on providing expert perspective on trends that drive regulatory enforcement activity. I'm Stephen Piepgrass, one of the hosts of the podcast and the leader of the firm's Regulatory Investigations, Strategy, and Enforcement practice group.

This podcast features insights from members of our practice group, including its nationally ranked State Attorneys General practice, as well as guest commentary from business leaders, regulatory experts, current and former government officials, and our Troutman Pepper colleagues. We cover a wide range of topics affecting businesses operating in heavily regulated areas.

Before we get started today, I want to remind all our listeners to visit and subscribe to our blog at regulatoryoversight.com, and the *Consumer Financial Services Law Monitor* blog so you can stay up to date on developments and changes in the regulatory landscape.

Today in our fourth episode, focused on the trending topic of artificial intelligence, I'm joined by my colleagues, Mike Yaghi, a partner in our group, as well as Chris Willis, co-leader of our Consumer Financial Services Regulatory practice group. We'll discuss the potential risks financial services companies face with increased reliance on AI, as well as the increased focus on AI by various regulators and state attorneys general. Mike and Chris, thank you for joining us today. I know this is a topic both of you are following closely, and I'm very much looking forward to our discussion.

Chris Willis:

Me too. Thanks for having me on.

Michael Yaghi:

I agree. It's great to be here with both of you. Thank you.

Stephen Piepgrass:

Well, Chris, why don't we kick things off with the question a lot of our listeners, I'm sure, have as well. Why are financial services companies interested in using machine learning algorithms?

Chris Willis:

Sure. And you can see there's been such an incredible uptake in the financial services space of the use of these models and it's because there is such a strong set of justifications for using them. As listeners have heard on the other episodes of this series of the podcast, machine learning algorithms allow for a much larger set of input variables to be assessed by a model, so you can consider more information on the front end. And then, they tend to make more precise and more accurate predictions based on considering those inputs, because they consider the interaction between variables, not just each variable standing alone, which is the way the old kind of models, logistic regression models worked.

And so, when you have access to a tool that can do that, it allows financial services companies to basically make smarter and better decisions about their lending activities and other aspects of their business, so that they can actually make money more efficiently, because they lend to more people who end up repaying, and make more accurate determinations about who's going to repay.

But in addition, because it allows the use of lots more data inputs on the front end, it allows creditors to consider data outside normal credit bureau attributes and thereby, make their credit models more inclusive and make credit available on a very accurate basis in terms of underwriting to a segment of the population that would never have had access to credit if all we were doing was just churning over the same credit bureau attributes again.

It's the desire to make more accurate decisions, but also the desire to make more inclusive decisions and use some of the very rich datasets that are out there, that go beyond traditional credit reports, that have really driven the very widespread adoption of machine learning models in the consumer credit industry.

Stephen Piepgrass:

That's a great point and a very interesting one because I know as Mike and I have seen, AI, the use of algorithms often gets a bad rap among regulators, and one of the things that they seem to focus on the most is the potential discriminatory impact, but as you just explained, these actually can be used for good in this space, to really extend opportunities to people that otherwise might not have them.

Chris Willis:

That's right, and we're going to hear, I think later in the episode, about some other good that can come from using machine learning technologies also on the discrimination front, because there are tools that you can use on machine learning algorithms that you couldn't use on the old kind of models, to maximize their fairness, but I'll save that for a little bit later in the podcast.

Stephen Piepgrass:

Can you talk a little bit about some of the use cases for the algorithms, particularly in the financial services space?

Chris Willis:

Sure. The thing that everybody thinks about first, and in fact, the thing that the consumer financial services industry thought about first was using them for underwriting decisions. For decades, 50 years probably, there have been automated underwriting strategies for different kinds of credit products, where people would take attributes off of somebody's credit report or their application, or both, and run it through what was used to be a logistic regression model. And make a decision about approving or declining someone for credit based on the outcome of that model, at least in whole or in part.

Taking those logistic regression models and replacing them with machine learning models was the first thing that happened in the consumer credit industry, and that's the area where we've seen the most widespread adoption of them. I'm not saying that they're universal, but there's a lot of them in the market now. Where those approved decline models for products that are decisioned in an automatic way, are now done by machine learning models.

But having had some initial success with using them in that context, you've seen a desire in the financial services industry to then use machine learning models in other contexts. And so, we

see for example, the use of machine learning models for things like servicing and collections, like looking at consumer behavior in response to various things that may happen to their account after they've already gotten the account. In terms of predicting when they might become delinquent or they might be having trouble repaying, to offer them opportunities to get on some kind of loss mitigation program.

Or, for accounts that are in collections, choosing the method of communication or the time of day where the person is most likely to be responsive. So, instead of calling them every few hours throughout the day, the machine learning algorithm can predict when you're most likely to answer the phone, so you only get one call instead of four or five calls.

Likewise, and unfortunately, fraud is a really big problem in consumer financial services. There are a lot of organized, very sophisticated criminals who perpetrate large scale fraud against consumer lenders. And so, detecting fraud and preventing it, and I'm talking about both third-party fraud, like you're engaged in identity theft, for example, or first-party fraud where you are who you say you are, but you're opening the account for the nefarious purpose of stealing from the creditor. Detecting that fraud by reference to a lot of the data that's out there is also something that people have started to use machine learning models too.

And then finally, we've seen the adoption of machine learning and artificial intelligence for customer interactions. A lot of us has probably had the experience of having a website or a mobile app with a financial services provider, where we can chat with a bot essentially, that's not a real person, and you can ask it questions and it can provide some information without the necessity of a human having to answer, unless you ask it a question it doesn't know. And so, there's been a significant adoption of that and there's machine learning behind those as well.

Those have been the use cases that I've seen so far, but I think we're likely to see it come into additional aspects of financial services in the future.

Michael Yaghi:

That's great, Chris, can you describe some of the regulatory risks associated with using this type of technology in the financial services space?

Chris Willis:

Sure. There's really, I would say, three of them that are foremost in people's minds from the regulator standpoint. First off, we have a long history of particularly the federal banking regulators like the OCC and the FDIC being very concerned with what we call in the industry, model risk management. The idea is that those federal banking regulators are charged with the responsibility of making sure that banks operate in a safe and sound manner. In other words, that they make loans in a way that's not likely to fail the bank and trigger a claim on the FDIC fund.

There's a large existing body of regulatory guidance about model risk management, to make sure that the model is actually accurate and predicts appropriately and will weather changes in the economy and continue to be predictive. Whenever you take your modeling technology and move and say, "Oh, I'm not going to use the thing we've been using for the last 50 years, we're going to do machine learning now." The regulators are concerned to make sure that their general guidance about model risk management is followed with respect to these new technology models, because they want to make sure that they are accurate today, in terms of allowing the bank to make loans that will mostly be repaid, and that they will be adaptable when economic conditions change. Like, we have a recession or unemployment goes up or whatever.

That's concern number one is, do they actually work and will they work not just in the short-term but in the long-term? And can they be updated and refreshed and validated the way the old

models were? That's number one. Number two is really the fair lending concern, which you made reference to earlier in the podcast, which is, you hear a lot of press about the capacity for a machine learning model to produce biased or discriminatory results, which is true, but it's no less true of a logistic regression model. You can build a logistic regression model off of people's credit bureau attributes and it will have a disparate impact almost every time. All credit models do. It's unfortunate, but true.

And there's no avoidability to it. If you underwrite consumer loans based on FICO score, that has a disparate impact. That doesn't make it illegal, by the way, right? No one's ever gotten into fair lending trouble for using FICO scores, even though they cause a disparate impact. And the reason for that is because business justification is a defense to disparate impact. And the business justification of accurately predicting who's likely to repay and who's not, will leave a practice legal even though it has a disparate impact. That's why FICO scores aren't illegal.

But consumer advocates and regulators are concerned that machine learning models can result in biased or discriminatory outcomes, even in a way that is illegal, that's not justified by business necessity. And so, I think they're very focused on reminding financial services providers that machine learning models are subject to the Equal Credit Opportunity Act, subject to these anti-discrimination laws, just like the old models are, and they need to be built and tested in a way that makes sure they are compliant with those laws.

That's regulatory concern number two, and it really is probably the one that occupies the most attention from regulators and from financial services companies to make sure that we, as the industry, are in a position to show that the model was built properly and that it can withstand a regulator's scrutiny on fair lending. So, that's number two.

The third regulatory concern, and this is the tail on the dog, I think, is one of the things that's unique about credit, particularly when you're talking about making a decision about whether to approve or decline somebody when they apply for credit, is the Equal Credit Opportunity Act requires a declined applicant to receive what's called an adverse action notice.

So, if you ever apply for credit and you don't get it, instead you're going to get a notice saying, "We declined you and here are the reasons essentially, of why your application was declined." The suspicion among some regulators and the CFPB particularly voiced this in a guidance document that was released in 2022, is that machine learning models can be so complex and so tied to the interaction between attributes rather than the operation of single attributes by themselves, that it may be difficult or impossible to explain what were the top factors that led to someone getting declined on their particular credit application.

My own view about this is that this was really a symptom of the earliest of the machine learning models that started to be adopted five years ago or something, 10 years ago, but I think the industry very rapidly grasped the need for explainability in machine learning models. And so, techniques arose pretty quickly that allow the accurate derivation of adverse action reasons from a machine learning model.

And if you ask me, I think the state-of-the-art method to do that is by using Shapley values, which I'm happy to tell you about, if you want to get into that. And that is the predominant method that we see industry use in deriving adverse action reasons from machine learning models is Shapley values. And so, I think that problem has been resolved. Shapley values, I think, are accurate and provide good basis for adverse action reasons. But I think the regulators are warning the industry that it's important, and so, we have to be in a position to prove that we're doing it and doing it correctly, but I think the technological challenge in doing that has really gone by the wayside.

Michael Yaghi:

Focusing on what regulators are concerned about, have the financial services regulators given any guidance to the industry on how to use machine learning algorithms?

Chris Willis:

Really, almost none, and that is a source of frustration for some in the industry. On the subject of just generally how to properly develop a machine learning model, to test it for fair lending, disparate impact, et cetera. I'm aware of little or no guidance from any of the regulators. I don't really think that exists.

My perception is the regulators are still learning about the models and they don't want to come out with something very specific or prescriptive when the adoption of the models is in a fairly new stage with the industry. And also, there's the opportunity for technologies and techniques to change rapidly. And so, we haven't seen any sort of general guidance about machine learning models.

And then on the third topic that I mentioned a moment ago, adverse action notices. It's kind of funny because in the summer of 2020, the CFPB published a blog post acknowledging that there is regulatory uncertainty and ambiguity about how you should derive adverse action reasons from a machine learning model. And by way of background, there's official commentary in Regulation B, which is the regulation under the Equal Credit Opportunity Act that provides specific methods for deriving adverse action reasons from logistic regression models. And it says, "If you do it this way, this is okay." Essentially.

Well, those examples don't really work for a machine learning model. And the CFPB published in July 2020, a blog post saying, "We understand that there's not guidance on this issue and it's the source of some uncertainty, and we're worried that uncertainty might deter the adoption of machine learning models because creditors aren't sure exactly how to do adverse action." Now, in fact, it did not deter the adoption of the machine learning models, because they're in the market now. But the CFPB said in that blog post, "We're thinking about revising the official commentary to Reg B, to provide this guidance."

So, I was excited. I thought maybe they were going to come out with that guidance. I mean, it's on a smaller issue, but still, that would be nice. Fast-forward to 2022, and the bureau puts out the bulletin that I was mentioning just a minute ago, that says in a very definite tone, "Creditors, you had better not use a machine learning model if you can't produce adverse action reasons out of it. And if you do, that's a violation of law. There's no exception for machine learning models." Which nobody ever thought there was. So, thank you, Captain Obvious.

But the bureau did not provide any of the guidance or clarity that they acknowledged was necessary and promised to provide in the 2020 blog post. That's an example of the federal regulators not just not giving guidance but acknowledging that they need to and then not doing it. That's symptomatic, I think, of the general lack of guidance on this issue. And so, everybody in the industry is forced to make their best judgment about the right way to do it. And we have a lot of ways to make the educated guess to do that, but we don't have a lot of experience or guidance from the regulators to empirically prove that whatever we decide is the correct way.

Stephen Piepgrass:

Interesting. And your observations in this area, I think are even more broadly applicable. Whenever there's new technology being developed and the regulators start to think about how to regulate it, how to address it, there is a significant risk of them stepping in too early and locking in the technology in the midst of its development, but before it can develop to its maturity.

And so, in some ways, I would think we should take comfort in the fact that regulators are hesitating, recognizing that this is changing on an almost everyday basis, when you read the headlines and you see all that's going on with the development in the AI space, I think in some ways maybe we should be glad that they haven't locked us in yet. But if history is any indicator of the future, I am sure that will be coming before we know it.

Chris Willis:

Yeah, we'll see. But one thing I'll say to the credit of the regulators is that certain of them are highly attentive to understanding the issues surrounding machine learning and are very educated about it and are consuming what's going on in the industry with a great deal of interest and analytical thought.

I see evidence of that, for example, with the CFPB, and so, I'm not sure that we're going to see the bureau do something reflexive and inflexible. They seem to be paying attention in a thoughtful way. They haven't done anything yet, from a guidance standpoint, but they definitely seem to be aware of the issues and asking the right questions.

Stephen Piepgrass:

Yeah, that's very encouraging. Chris, as we wind toward the end of our conversation today, I'm sure our listeners would love to hear your perspective on best practices, especially in an area where we don't have a lot of guidance yet. What should they be thinking about in terms of best practices as they're developing and adopting machine learning models?

Chris Willis:

I have a set of best practices that I use in advising clients on this, and I'll give the high level of it. These have been developed really in reaction to public commentary by consumer advocacy groups, my own conversations with them, and my conversations with regulators. So, even though they represent my best judgment, I think it's a best judgment that I've at least had the opportunity to socialize with some pertinent audiences. And so, I have a pretty good feeling about what I'm about to tell you.

First things first, when you develop a machine learning model, the way that works is you have a set of data to train it on. The types of machine learning models that the financial services industry use is basically, you'll take a big batch of data and say, "Okay, here's a batch of loans. Here's the information we had at the time the loans were originated, and then here's the outcomes." And then you throw that in and you let the machine learning model ascertain the correlations between the input variables and the combinations of the variables, and the outcome that you're trying to predict, which is usually, did somebody pay or not?

One of the criticisms of machine learning models, particularly from a biased standpoint, is that if you have a training dataset or a development dataset, those are the same thing, that doesn't have a broad or diverse enough population in it, that if there are members of protected classes in there, and the machine learning model won't know that, of course, it doesn't know what people's race or gender or age is. But if you have underrepresented groups there, the machine learning model can focus on idiosyncrasies about that group and will appear to correlate them with outcomes in a way that wouldn't be borne out in a larger population. We call that overfitting a model or the model overfitting itself.

And so, one of the potential sources for bias in a model is overfit on characteristics that are common to members of protected classes, when the development dataset isn't diverse enough or large enough or broad enough. So, best practice number one is, let's start with the most diverse, broad set of development data that we can get. Now, there are limitations to that. You

don't just walk into the store and buy a set of development data. People don't sell those a lot and they have to have certain characteristics and be similar or from your product. But when you have the choice, choose for more inclusion in your development dataset. So, that's number one.

Number two, think about your input variables before you start the training process, because so much fair lending risk can come from just a review of the attributes in a model. And certain attributes are things that I would look at or a regulator would look at and wouldn't look twice about. Things like your conventional credit bureau, like, "How many delinquent accounts do you have?" Or, "When was your most recent bankruptcy?" Or, "What's your credit utilization." Or, stuff like that. "What's your debt to income ratio?"

You can put those in a model all day long and it'll never raise an eyebrow, but there's a lot of data sources available now that go beyond those traditional, classic credit bureau type of attributes, and some of them will wander into either being or being close proxies for being a member of a protected class. And we know that regulators have certain attributes that they are highly suspicious of from a fair lending standpoint. Do yourself a favor and look at the attributes before you start training the model. And if there's something that looks like it's really going to be problematic, let's leave it on the cutting room floor early, so we don't get halfway down the model build process and then have to extract it out, because we determine it's a fair lending risk. So, that's number two, variable selection.

Number three. Usually, we will have variables in a model that have a bit of concern around them because they could cause disparate impact, but they also may be legitimate if they have enough business justification associated with them. And always in these situations, we have to weigh business justification versus regulatory risk, because again, business justification goes to how much an attribute contributes to the model's predictive performance, which we care a lot about from a business standpoint. But also remember, business justification is a defense to a disparate impact claim.

So, we have these variables that we're not sure about, that will be in contention for being in the model. Well, if we have ones that we're a little bit worried about, another thing we can do before we start the training exercise is just do some univariate analysis. Just look at that variable by itself and say, "How strong is the correlation between this attribute and being a member of a protected class? Is it a very high correlation, such that we would worry that it's a proxy for actually putting protected class in our model? Or, is the correlation very weak?"

And the thing is, if we do that test before we build the model, it gives us great information. If we find a proxy, we know not to use it. If we find that it's not a proxy and someone later challenges the variable, we're in a position to say, "Oh, no, it's not. It's not a proxy. We tested it and here's what the testing showed. And we can show that it's not correlated with being in a protected class." So, doing that univariate testing is candy. It's not necessary, but it's a best practice. So, that's number three.

Number four is, when we build the model, now we're going to actually train the model, we want to pay particular attention to making sure that we document the business justification of the model. This isn't a hard thing for modelers to do. This is like asking them to breathe, because their whole function in life is to build the best, most predictive, most accurate model that they can. Right? That's what they do for a living. There's no problem getting them to do this. It's just a question of making sure the documentation shows a thorough, rigorous process that is very empirically based to show that, "Hey, this model performs really well." Not just in absolute terms, but in comparison to other alternatives like using a FICO score or a VantageScore, or our old model, or something like that. So, we like to see the model documentation show both of those things. So, that's the next best practice.

The next one is, I think in today's environment, it's necessary to do fair lending testing on a credit model. What that means is testing the model output for disparate impact, as well as obviously looking at the business justification of the model, as we just did in the prior step. But it also means looking at individual attributes and assessing, how much do they contribute to the predictive accuracy of the model, versus, how much do they contribute to disparate impact in the model results?

And where you have attributes that contribute little or nothing to the predictive accuracy of the model, but contribute a lot to disparate impact, those are candidates for removal from the model. Likewise, if you have attributes that are highly predictive and help the model a lot, you know you're not going to take them out, even if they do cause some disparate impact, because again, business justification is a defense to disparate impact. That's the kind of testing that we've always done with logistic regression models. The method just has to be different for a machine learning model, because the old methods that we used for logistic regression models just don't work in the machine learning context. So, you just have to test it a different way, but you're basically doing the same analysis.

And then the final best practice is one that I teased the audience with at the beginning of the podcast. It's about de-biasing models. The thing is, when you use a machine learning model, there are technologies in the marketplace today that will allow the model to generate various versions of itself by tweaking aspects of how the model behaves. I won't get more technical than that for the purpose of this discussion, but it basically allows an exercise where you can build five or 10 different versions of the model easily and quickly. They're not major differences, they're just tweaks on how the model handles data and returns results.

You can then do fair lending testing on the disparate impact of each of those variations of the model, and it puts the creditor in a position to see both the predictive accuracy and the level of disparate impact from every version of the model. Well, what that allows the creditor to do is to make sure that they can have the option of picking a version of the model that has the least amount of disparate impact to still achieve the level of predictive accuracy that they need. Then they are in a position to say, mathematically, "I picked the least discriminatory alternative that still fits my business justification." Because remember, that's the third part of the classic disparate impact test, right?

First, is there a disparate impact? If so, second, is there a business justification? If there's a business justification, you're okay. But third, if the plaintiff or the government proves that you could serve that business justification with a less discriminatory alternative, you're still on the hook for disparate impact liability. And so, what you're doing with this de-biasing technology that's on the market now, is you're putting yourself in a position to prove that you adopted the least discriminatory alternative.

The regulators are very aware that this technology exists in the market, and so, I think the failure to use one of these techniques will get them asking, "Well, how do you know this is the least discriminatory alternative? And therefore, why shouldn't I be suspicious of your model?" So, I think that's my last best practice is to use these de-biasing technologies, which are unique to machine learning. You can't do this easily with a logistic regression model, but you can do it with a machine learning model. So, it's another way that the technology is actually superior to what we had in the past, not just in terms of predictive accuracy, but in the ability to generate a model that is the fairest model possible. I apologize for the long-winded answer, but those are my thoughts about best practices for machine learning models in the financial services industry.

Stephen Piepgrass:

Well, thank you, Chris. And I know our listeners really appreciate the practical advice on the best practices, and I love that we came full circle talking there at the end about how in fact, machine learning can be really good for consumers and from a regulatory perspective, actually beneficial in avoiding discrimination. Really appreciate your insight and expertise, and thank you for joining us.

Thank you, Mike, as well, for sharing the co-host responsibilities with me today. I know our listeners very much enjoyed the insights. And I want to thank the audience for tuning in. As always, we appreciate you listening and don't hesitate to reach out to us, the Troutman Pepper team, if we can help you in any way. And Chris with our Consumer Financial Services Regulatory group.

I hope you'll join us for our final AI podcast episode where we will be discussing AI's impact on background screening. Please make sure you subscribe to the podcast as well as our Consumer Finance podcast using Apple Podcast, Google Play, Stitcher, or whatever platform you choose. We look forward to having you with us next time.