***REGULATORY OVERSIGHT* PODCAST — AI: TECHNOLOGY, OPPORTUNITIES, RISKS, AND BEST PRACTICES (PART TWO)**
HOST: STEPHEN PIEPGRASS
GUESTS: JIM KOENIG AND CHRIS WILLIS

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 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 so you can stay up to date on developments and changes in the regulatory landscape.

Today's episode is the second in our series on artificial intelligence. AI has captured the imagination and generated excitement across businesses, but at the same time, developments in AI have also raised public concerns and spawned regulation that sometimes threatens to outpace the technological innovation we're seeing in that area.

Today I'm joined by my colleagues Jim Koenig, who co-leads our Privacy and Cyber Practice Group and Chris Willis, who co-leads our Consumer Financial Services Regulatory Practice Group. You can follow the work of Jim and his team by subscribing to their *More Privacy, Please* newsletter, and you can follow Chris and his team through their *Consumer Finance Podcast* and their blog, the *Consumer Financial Services Law Monitor*.

Our discussion today will provide a little background on AI, including the opportunities and risks it presents, as well as emerging global best practices surrounding the collection, use and sharing of data and its use in AI. Jim and Chris, thank you for joining us today. I know this is a topic you're both well versed in and I'm very much looking forward to today's discussion.

Chris Willis:

Us too. Thanks for having us on, Stephen.

Jim Koenig:

Great to be here. Thanks for having us.

Stephen Piepgrass:

Absolutely. Jim, maybe you could kick us off with just a little background on what is AI.

Jim Koenig:

This is the actual real Jim, not an artificial intelligence inspired Jim providing the response here. Technically, AI's been around in various forms for a long time. It's the simulation of human

intelligence processes by machines, especially computers. Examples include applied AI to expert systems, natural language processing, speech recognition, machine vision, all sorts of things – but a little technical. How does it work? One example, the one that people when they think about AI, think about, is predictive AI. Predictive AI systems work by ingesting large amounts of data, training the algorithm to look for correlations and patterns, and using these patterns to make predictions about future states. So what does that really mean? An example can be found that by looking at patterns, there's some research out there that says loan requests written by defaulting borrowers are more likely to include words relating to their family, mention of God, borrowers financial and general hardship or short-term focus words. Patterns that human beings may not be able to perceive that might be able to provide some intelligence.

Maybe it's just noise in the system, but if you ask an artificial intelligence researcher, he or she would say that it's a set of algorithms that can produce results without having to be explicitly instructed to do so. It's the simulation of natural intelligence in machines that are programmed to learn and mimic the actions of humans, and these machines are able to learn with experience and perform human-like tasks. As AI continues to grow, it'll have a great impact on the quality of our life in many different ways.

A way to think about it, and there's lots of frameworks that people talk about different aspects of AI used in different industries and skills, but if you distill them all down, there's really three cognitive skills or principles. There's learning, where it acquires data, creating rules or algorithms, but it's repeating a human task. Raising or lowering a shade or using voice to mimic otherwise human instructions like Google or Siri. The next level is where the technology starts to have reasoning and in a reasoning type, the machine is choosing the right rule or algorithms to reach a desired outcome. So maybe approving a loan based on certain metrics so it can do it on the fly online. And then the last one is self-correction, where the model actually and the algorithm teaches itself and improves outcomes and efficiency by continually fine-tuning the algorithm, ensuring that it provides the most accurate results possible. An example of that, just simple learning the best path for an autonomous vehicle to take to work or which lane to be in for traffic in a connected car. So those are some examples.

Stephen Piepgrass:

Thanks, Jim. That's a great primer and background on this subject. Chris, I know you've got a lot of clients in the consumer financial services space. Tell us a little bit about why it is that clients are interested in using AI.

Chris Willis:

Sure. The basic reason is there is an incredibly strong business rationale behind the adoption of machine learning and artificial intelligence, both in my industry, financial services and elsewhere. Because as Jim was talking about, machine learning and artificial intelligence algorithms take advantage of moderate advances in computing power to allow the machine learning algorithm or the AI to assess a hugely greater amount of information than old types of algorithms like logistic regression models were capable of doing. And so that creates the advantage of being able to allow businesses to make faster, more accurate decisions than were possible before under either a manual process or under the old types of algorithms. And it allows them to make those decisions in a very objective way, assuming that we build the

models accurately and then it creates the capability for machines to do things that were never possible before. For example, using AI in healthcare as part of diagnostic tools.

You read in the media all the time about how a machine learning algorithm can predict someone having cancer years in advance of a doctor being able to detect it under conventional methods. In financial services, as Jim mentioned, you have the use of machine learning algorithms to make underwriting decisions, but it's also for detecting and combating fraud in a way that recognizes the patterns of fraudulent activity, particularly by organized fraud rings. In cyber defense and security, which is Jim's area but I'll say a little bit about it. It allows people to detect and identify malicious materials and detect cyber-attacks in advance of them actually causing a cyber incident. And so the point is this is a technology that has a massive amount of appeal, and so it's no wonder that companies across the economy are interested in adopting it.

Our job as lawyers is to make sure that we guide our clients to do it in an appropriate way that allows our clients to realize the benefits of the technology and avoid some of the risks and criticisms that are out there in the media and among regulators and public interest groups against the use of the technology. And that's really what we're all about here at Troutman Pepper.

Stephen Piepgrass:

Thanks Chris. Jim, it may be helpful for our listeners to think about AI in terms of types or categories. Could you talk a little bit about that?

Jim Koenig:

Sure. There are different types of AI. There's machine learning which teaches a machine how to make inferences and decisions based on past experiences, identifies patterns and analyses based on past data to infer the meaning of those data points. Deep learning. Deep learning is a machine learning technique. It teaches a machine to process inputs through layers in order to classify, infer and predict the outcome. The differences between that and another type is neural networks. Neural networks work on similar principles to human neural cells. There are a series of algorithms that capture the relationship between various underlying variables and process the data as a human brain does. And there's natural language processing, which is a science of reading and understanding and interpreting language by a machine, and computer vision or algorithms that try to understand an image by breaking it down and studying the different parts of the object. And this helps the machine classify again and learn the different types of images.

There are more, and there's different ways to classify them. But a really important way for the non-AI researcher or scientists to think about it, as I mentioned before, is really about whether you're just repeating human tasks or you're trying to create some reasoning and replace human decisions or the machine is learning from itself and trying to have human behavior in it. These are the different types of AI applied in different industries in different ways. Not to get too technical.

Stephen Piepgrass:

Maybe you could also talk a little bit about, and this is moving us toward the regulatory framework that I know many of our clients and listeners tuned in to hear about. What are some

of the risks associated with AI, particularly if treated as the classic black box? Can you speak a little bit to that?

Jim Koenig:

The risks that are there are simple but diverse. The simplest is that the machine or the algorithm produces a wrong result or an unfair result or something that's unsafe. And it's very difficult to be able to see around the corner when you're building these models to see where the decisions would go and what type of information could set it off course and what type of risk and harms there are. There are a number of emerging frameworks. One of the ones that I think clearly spells it out is the NIST framework. On January 26th of this year, NIST released an AI risk management framework along with the companion set of other documents and playbooks that'll be very helpful for companies and researchers alike to be able to think and look around the corner about what type of harms there are. Now, very often we think of, and NIST described them, AI risks and trustworthiness.

All that computer processing is, is a replacement for some human decision making or human task, and how trustworthy is it? And NIST breaks it down into seven different types of risk. First, will the decision actually be valid or reliable, just plain correct. Next is safe. For example, if you're making decisions about AI, an autonomous vehicle and the decisions are incorrect and it ends up hitting a person or a truck or another vehicle, the result there would be unsafe. Another harm is that the result at the end, there's accountability and transparency. Making sure that the result was based on good principles, good math, and I'll talk about bias in a moment, and that finally the process is privacy protected. And very often we think about the harms in AI in terms of consent and people give permission. This framework that NIST put out actually signals a shift from rather than just consent, people can give permission.

The fact that we're looking at frameworks and regulators are starting to look at frameworks in terms of harm, companies and programmers have to protect individuals from themselves, using a harm-based framework rather than consent and privacy's right at the center of it. And last but not least, is definitely that the harm can be that it's unfair or that it incorporates aspects of bias, which has been a part of the discussion about AI and making sure that it's fair, inclusive, and makes sure that it provides a benefit for all and doesn't produce one of the harms that we discussed.

Stephen Piepgrass:

Chris, and maybe you can speak a little bit to this bias issue. I know in the AG space that our group specializes in, that's something AGs are looking at very, very closely. And I know the consumer financial services world in particular has drawn a lot of scrutiny in that maybe you can provide your perspective on this, the risks of bias.

Chris Willis:

Thanks, Stephen. And you're right, the possibility of bias or discrimination in machine learning or AI models is in the media all the time, and it's top of mind for both state attorneys general and the financial services regulators like the CFPB or the Federal Trade Commission. I think it's important to recognize that a machine learning model can have bias as we think about it legally. And I think the easiest way to think about it legally is using the decades old framework that we

have under employment and fair lending laws of disparate impact. That is that it disproportionately impacts someone who's in a protected class and lacks a business justification or doesn't have a less discriminatory alternative that'll achieve the same business justification. That's of course how I think of bias because bias by itself isn't illegal, disparate impact is under, for example, Title VII or the Equal Credit Opportunity Act.

And so I think it really does us a lot of good to think about where can bias come from, and I'm thinking about illegal bias in a machine learning model. So I've got a few things to share with the audience about what to look out for so that when you're starting to build a model, you're attuned to these issues. First off, as Jim said before, the way machine learning models are built is you start with a development or training data set. That's the data set that the algorithm is going to use to recognize patterns and draw correlations between input variables and the outcome that's trying to be predicted or achieved. One of the criticisms of machine learning model processes can be that the development data set isn't big enough and specifically isn't inclusive enough of members of protected classes.

So if you only have a few members of a particular protected class in the development data set, the tendency is for the model to do what's called overfitting. Which means drawing false conclusions and false correlations based on spurious relationships between inputs and outcomes and will basically judge the members of those protected classes unfairly and inaccurately because of the thin data associated with them in the development data set. So making sure we have a good development data set is really important.

Second is looking at candidate variables. One of the natural reactions of a lot of model builders is to say, "Well, hey, I have 5,000 attributes available to me, so I'm going to use all of them because I don't know where my predictive attributes will come from. I don't know which ones will be better, so I'm going to let the machine sort it out." And that's fine if all you're thinking about is building a model, but if you're a little bit external focused, you can say, "Well, hey, let me look at those 5,000 and see if there's anything that looks discriminatory or looks like it could create bias like someone's race or gender or religion or things like that." Those are pieces of data that honestly should not be in most models because they directly have the capability to introduce bias because you're going to let the model learn that someone of this race or this gender or this religion acts in a certain way and then the model is acting on a stereotype that's unacceptable to us as a society.

Third, sometimes a model can be built without a lot of insight or control over the training of the model. In other words, they just let the model do its thing and reach its conclusions and build itself without understanding exactly how it's working and why. And that lack of supervision over the training process can allow bias to come in without the model builders even knowing anything about it. And that goes hand in hand with the next thing that I wanted to mention, which is if there's an inattention pay to bias or disparities in the model output. In the lending area, consumer financial services where I work a lot, it's very important for creditors to test models to make sure they don't have a disparate impact. That involves an analytics of the output of the model to see how different people fare under the model. Well, if you do that exercise, then you're going to know. If you don't do it, then you won't know. And so the point is people sometimes don't do it.

And then finally, remember the classic formulation of disparate impact as given to us by the Supreme Court in the 1970s was that even if a facially neutral policy or practice has a disparate impact, it's okay if it has a business justification but not okay if that same business justification

can be served relatively equally well by a less discriminatory alternative. And so a lot of times the model process is called at an end when the model is built. Well, okay, we're done. We've got it. But the point is, if you don't consider whether there's another way of formulating the model that might reduce bias or reduce disparities, but get you to the same business result, the job's not done yet. What I'm saying is you don't call the job done until you consider those alternative formulations or versions of the model. Those are all areas where bias can sneak into a model without honestly people really being conscious of it, but nevertheless, the model can have a disparate impact that can create a problem for the company using it.

Stephen Piepgrass:

Those are excellent points. Hadn't thought a lot about a lot of those that you raised. I mean, I love the point you made about you know sometimes it's not so much what is the data that's in the data set. It's also about, especially when it comes to bias, what data should be left out of the data set? Excellent points there. One of the things our listeners tune into and I think most look forward to is our crystal ball. We pride ourselves in keeping track of regulatory developments. And because of that, and I know this is something both of you do in your groups and your practices as well, we're able to look a little bit into the future and maybe it would be helpful to hear from both of you what you see in the coming year when it comes to the regulation of AI.

Jim Koenig:

I'll take it first. I think with respect to AI, the here and now, I don't even have to look deep into the crystal ball. But the here and now, some of the initial attempts at regulating AI is through privacy. For example, California and the CCPA and Virginia and its new Data Protection Act, Colorado and Connecticut, they have provisions in them that have restrictions or disclosure obligations or different things around automated decision-making technology. Said differently using AI to make decisions or to produce results, but it's not checked or somehow quality reviewed by humans. And so you could end up with a machine or an algorithm running producing unfair, biased or improper or wrong or unlawful decisions. These laws say, "We don't know what can go wrong, but we'd like a human to check it. Or if not, there has to be a series of disclosures."

California and Virginia are already in effect while the other laws come into effect July one this year or the beginning of next year. Also, one more trend that came at it was in DC. In DC, they went right after sort of using privacy biased algorithm. So there's a proposal to Stop Discrimination by Algorithms Act that was proposed in 2021 that makes it illegal for companies to use algorithms that make eligibility determinations based on a series of protected class and other unfair situations. And not only is it regulatory enforceable, there would be potentially a private right of action. And so down the road, whether it's through privacy or directly focusing on bias, those things are bubbling up and companies should prepare to make sure that their AI activities are well thought through, documented, and designed to avoid harm.

Chris Willis:

And Stephen, if you don't mind, let me just pick up where Jim left off. The last thing Jim said is the most important message for our audience. When you use a machine learning or AI process, it's really important to thoroughly document the process by which you built the model, tested it,

made sure to avoid bias and discrimination, et cetera. Because going back to your question, what do we see for the future? From a regulatory standpoint, I think there's a significant amount of suspicion about machine learning models and the harms that they can do. I think regulators are going to come at this with the presumption that the machine learning model needs to be proven appropriate, that we are not going to assume that it's appropriate. And at the same time also, I think there's going to be a desire to show the market an example of a machine learning process that wasn't appropriately managed, as the subject of an enforcement action.

The regulators, I think, right now are not in a position to be all that detailed and prescriptive about exactly what you must or must not do in the details of developing a machine learning model. But they are definitely to the point where if you just don't do anything or fail to document the model and test it for discrimination, et cetera. They're ready to use that as an example. And they're looking for that example to send the message to the market, that Jim just delivered in his comments, which I agree with about making sure you develop the models the right way. And so that sort of sword is hanging above the head of every industry that uses machine learning. And I think we all need to be aware that we need to get out from under that sword and that we can do so by appropriate model development and governance techniques.

Stephen Piepgrass:

Chris, that's a great point, and it's something we talk about a lot that regulators are very aware that they are at the cutting edge of the development of law. It's often called the law of settlements. May not be codified in court opinions, but it is in consent orders. And when it comes to areas where technology is outpacing legislation, or even the courts, the first to tackle those issues tend to be the regulators and they are on the lookout, especially in this area of AI to create that law. Those who are using it, just be very aware of that. That really, I think, segues well into the last component of our podcast today, which is any final tips or best practices that our clients should be considering if they're thinking about using AI or already are using AI in their processes?

Chris Willis:

Absolutely. And so Jim and I put our heads together, we came up with a top 10 list of best practices, which we're about to share with the audience, and I'll do the first part of it and then I'll let Jim take over. These flow from the things we've already talked about during the podcast today.

So number one, when you're building an algorithm, get as broad and inclusive a set of training data as you can get your hands on so that the model fairly represents everybody in the population and doesn't create idiosyncratic connections that aren't real with members of small groups that might be underrepresented in the data.

As we were talking about before and as you made the point, Stephen. Number two, think about what data we should allow the model to train on, it's not a question of what we can do. It's a question of what we should do. Take a red pen and mark out the ones that you don't want on the front page of the newspaper and just don't let them even enter the model development and training process.

Third, if you have attributes in a model that look like they might be helpful but also look like they might create some risk of bias or discrimination, you can look at those variables in isolation prior to even including them in the model build process. Just do a single variable correlation to see, "Hey, how correlated is this attribute with being, say, a member of a particular race or gender or religion or whatever?" The outcome of that will give you information. Either you will know that it's a proxy for that protected characteristic and so I shouldn't use it. Or if you're later challenged for using it, you can say, "Hey, I looked at it and it's not highly correlated with being a member of a protected class." It gives you a great way forward, honestly, regardless of how it comes out.

Next, number four on the list is be sure to thoroughly document the business justification of the model. Business justification is the key to defending against any claim of disparate impact. And so proving that the model is effective and accurate and is superior to other alternatives, like using something off the shelf or using the old way that we did it or using a judgemental process is key to showing that the model serves that business justification.

Of course, depending on the industry you're in, number five would be considered doing disparate impact testing of the model output, just as I was talking about earlier. Either for reputational reasons or for legal reasons, you may want to be in a position to say, "I've tested the model and it doesn't have significant bias." You won't know that unless you do that kind of disparate impact testing.

And then number six is one of the most interesting things about machine learning and artificial intelligence is the advent of automated de-biasing technologies. That will sit there and take a model's output and make tweaks or fine-tuning to the model to allow it to reach the same or almost the same level of business effectiveness, but sometimes at a dramatically reduced bias against members of protected classes. We call this de-biasing, and there's several examples of that technology that are on the market. Investigate and use those because it allows you, again, to defend the model and say, "Look, I made the best model I can make. It has the most business justification and the least amount of bias or discrimination that is possible, and I can mathematically prove it." So those are numbers one through six on the best practices, but Jim's got some more for you. So Jim, go ahead.

Jim Koenig:

Number seven on the list. An emerging best practice that many companies are starting to undertake are to conduct ethics assessments. To identify discriminatory impacts and privacy implications such as the identifiability or small data sets, where you actually could derive specific individuals who are in your dataset or a group of individuals, which then depending on the individuals and their traits, could lead to bias or unfairness.

Number eight on the list, do a lot of work with companies in front of the FTC, state attorney generals, the OCR in healthcare. And here, a lot of times when there have been questions about their practices and using AI or machine developed entrepreneurial technologies. They've gone ahead and established an internal ethical data collection, use, and sharing charter to spell out the key principles of what they're doing. As we talked about before, there's a patchwork of laws evolving not just at the US state level where they're taking consent or automated machine decision making versus biased algorithm approaches. Even in Europe or following in Brazil, they're taking a harm-based approach, outlawing certain types of business processes and uses that they think are improper using the AI against individuals. And so there's a patchwork of

laws, having a charter allows you to have a center. That you can use to be able to develop techniques and practices that will allow you to have a global approach to AI as opposed to AI algorithms and systems that are growing up based on the different laws and different jurisdictions.

Nine, developing contractual requirements in minimum data handling and security controls for the vendors and third parties with whom data is shared and making sure that you pass along these ethical requirements. So if you're a company, you want to make sure that your vendors who are doing the algorithm development and training, make sure that they maintain good security over the data. They don't reuse it for other purposes. Or if you are one of the vendors and you're helping create algorithms and may be using it across multiple clients and the data sets that you're using may have come with restrictions or you may want to add them to make sure that your customers don't use that information for discriminatory purposes or for targeted advertising or to track individuals' locations. There's a lot of AI in those areas that are used for marketing and new purposes, but increasingly they're being viewed by the FTC and other regulators for improper and unfair uses of those technologies based on location, targeted advertising, and just discrimination across the board.

And finally, number 10, most important of all, pick a good name for your AI system. Whatever you do, don't call it Skynet from the Terminator or Hal 9000 from 2001 Space Odyssey, that ultimately turned against humans. Maybe pick a good name like JARVIS from Marvel's Iron Man or Cerebro from the X-Men.

Chris Willis:

So Jim, we shouldn't use VIKI from I, Robot either, is that what you're saying?

Jim Koenig:

Exactly. Well, I think if you document your AI processes and explain it, you'll have good facts and a good name. That's our top 10 list of the emerging best practices that companies are doing to be able to build their systems, to be able to anticipate and comply with the current laws that are out there in trying to look around the corner as they develop these new AI systems and innovation technologies that are designed to improve our life, make things easier, improve safety, and all the things that we've shared here today.

Stephen Piepgrass:

Well, thank you Jim and Chris, love the top 10 list, and I really appreciated the insightful commentary and conversation with both of you. Today I know our listeners appreciated your insights as well and learning about what their companies can do to avoid regulatory pitfalls when it comes to the use of AI.

I want to thank our audience for tuning in today. As always, we appreciate you listening and don't hesitate to reach out to the Troutman Pepper team if we can help you in any way. I hope you'll join us for our third AI podcast episode where we will be discussing AI's impact on the Healthcare industry. And please make sure to subscribe to this podcast as well as Chris's *Consumer Finance* and Jim's *Unauthorized Access* podcasts through Apple Podcasts, Google Play, Stitcher, or whatever other platform you use. We look forward to having you join us next time.