THE GOOD BOT s01e07: AI Discrimination and Emerging Best Practices (Pt. 2)
Recorded 11/07/23

Brett Mason:

Welcome to the Good Bot, a podcast focusing on the intersection of artificial intelligence, healthcare and the law. I'm Brett Mason, your host. As a trial lawyer at Troutman Pepper, my primary focus is on litigating and trying cases for life sciences and healthcare companies. However, as a self-proclaimed tech enthusiast, I am also deeply fascinated by the role of technology in advancing the healthcare industry. Our mission with this podcast is to equip you with a comprehensive understanding of artificial intelligence technology, its current and potential future applications in healthcare and the legal implications of integrating this technology into the healthcare sector.

I'm excited about this episode today, which is part two of our series, focusing on AI bias and discrimination and emerging best practices. In this part two, we highlight a fireside chat with industry leader, Pedro Pavón. Pedro Pavón is the Global Privacy Director of Monetization, Privacy and Fairness at Meta. Pedro and his team advise Meta executives on privacy and fairness matters related to ads and monetization across all of the company's platforms.

Prior to joining Meta, he served as the Director of Senior Privacy Council at Salesforce among other position. He's joined today in this fireside chat with three of my colleagues from Troutman Pepper, who you heard from in part one of this series. Jim Koenig, Alison Grounds, and Chris Willis. I hope you enjoy this chat with one of the leaders in the technological sector from Meta, Pedro Pavón.

Jim Koenig:

Well, welcome everybody. I want to welcome you to the next in the Troutman Pepper series, Managing AI: Risk, Reward, and Regulatory. Today's session we'll focus on AI discrimination and emerging best practices to avoid bias and discrimination. We have Pedro Pavón from Meta. Pedro is a long-time friend, we're lucky to have Pedro today. Pedro is a leader in the Meta privacy program. He has been at a number of other leading technology companies through his career, and his focus now is on privacy, data monetization, and fairness at Meta. And so we're going to have a chance to talk to Pedro and learn insight from his experience about best practices and avoiding bias and discrimination.

Pedro, thank you for agreeing to have a discussion with us. But before we get into deep into the substance, love to hear a little bit, for the folks that are participating, to hear a little bit more about you. You're a thought leader, you're a privacy professional, a social media influencer, and you've had a fascinating career. Can you share with us a little bit about your background and your history that's led to your current role at Meta?

Pedro Pavón:

Thanks, Jim, and thanks a lot for inviting me. One of the things I talk about quite often is sort of like exclusion in the AI conversation of voices that are people that are members of groups that traditionally don't have a leading role in how we develop plans and methodologies around addressing new challenges and I'm really grateful to be here and I appreciate you bringing me along, so thanks a lot. Social media influencer, I don't think I qualify, but I do post often on LinkedIn and on Threads, Meta's newest social media platform, about AI and about fairness and discrimination in the context of AI. So if you're interested in hearing my musings or participating in those conversations, they're always ongoing on LinkedIn. Just hit a follow there, hit a follow on Threads. And I'm super engagement oriented, so if you tag me or mention me in something, I will reply.

My career and how I got here is sort of an interesting tale, but I'll be really brief. I went to law school in the late 2000s and I was lucky enough to be a research assistant with a professor who was leading the Cyber Security Legal Research Center at the University of Florida, and I started to become really interested in the intersection between data privacy and... Excuse me, between privacy and technology, what we now call data protection and data privacy and my previous service in the military in an intelligence capacity also sort of all commingled into a really powerful interest and curiosity about this topic. And so that's really where this all springs from, is just being able to sort of have a free-range law student mind about privacy issues during law school. And that developed into passion and my practice. I've worked on sort of all three sides of the law when it comes to privacy.

I've been in the regulatory side, I've been at the law firm, and now I've been in-house for the last 10 years at some pretty big tech companies. I'm at Meta now, I was at Salesforce before, and

Oracle before that. So I've got this wide spectrum of exposure and experience that hopefully makes me capable and competent at giving advice around some of the new challenges we're facing. At Meta, I know I'm going to preempt the next question because I think the next question is about what do I do at Meta, so I'll just answer it. I've been in traditional legal roles up until I joined Meta.

At Meta, I'm in sort of a hybrid role, which is I lead a public policy and internal policy team, and our job is sort of like twofold. We do sort of your traditional public policy engagement with regulators, policy stakeholders, government officials to get a better understanding of what external expectations are with regard to Meta's monetization practices. And that's our ad stack, basically anything we sell that you can't touch or anything we monetize that you can't touch. My team's responsible for the public policy, external engagement on those surfaces.

And then internally, we do a lot of product advocacy work with our internal stakeholders, engineering teams, sales teams, go-to-market teams to make sure that we create the most harmony, I use that word, convergence, the most convergence possible between external expectations and what Meta is actually bringing to market as features and products. What I love about this is that I get to maintain a very strong and direct connection to the legal components of building and designing products because I'm in a public policy capacity and I partner very closely with our legal team, and because I'm a lawyer, that partnership is much easier to manage and advance.

And then I also get the benefit because I'm not in this rigid legal role to not have to limit my contributions to what's legal, what's not, what's lawful, what's not, what's the definition say and what does that mean, but instead get to look around the corner and try to anticipate future threats and start to make sure that we start building durability by design, through privacy features and guardrails to ensure that we're not just complying with current or immediate requirements, but we're building products and services that will benefit users and be durable and defensible and stable for the long term. Because if we can do that, that makes the need for increased regulation less, it makes our user experience better and it makes, obviously, as a for-profit company, our ability to generate revenue more stable. So I hope that's an explanation of the job and who I think I am at least in the profession.

Jim Koenig:

No, that's really great. Hey, tell us a little bit about AI at Meta and the types of projects or things that you've been involved in to the extent that you can talk about them publicly.

Pedro Pavón:

Yeah, absolutely. Look, I think with the OpenAI, ChatGPT moment and Midjourney and DALL-E moment in the last 12 to 18 months or so, I think there's begun a public common dialogue about AI. And there's a little bit of a myth out there that OpenAI is sort of at the forefront alone on the development of AI, but we know that that's not true. Companies like Amazon and Google and Meta and Apple and others have been investing heavily in not just generative AI development, but AI development across all sorts of surfaces and use cases for many, many years. As far back as 2014, I created... Was it 2014, yeah, 2014 or 2015, I created the sort of AI governance committee at Oracle. This is how far ahead we were there, that's eight years ago, and Oracle was already implementing guardrails that now are considered standard and required. Well, we were doing that in 2015, 2014 back then at Oracle. Same thing for Salesforce, when I was there, we were way out ahead and when I joined Meta three years ago, the company was already very, very heavily investing in all of this.

What does this mean for people and products? I hope most of you have heard of Llama 2. Llama 2 is Meta's sort of widely hailed open source large language model that's trained on trillions of tokens of data and fine tuned by over 1 million human annotations and it's open source. So it means we made it of unlike some of the more closed models like OpenAI's or other companies, we decided to turn ours over so that we could have distributed innovation on our model and create really interesting cool surfaces for people to design new technologies on. And so that's sort of a different way for a large company to roll out large language models and large learning models and we'll see how that plays out.

But I'm really optimistic that it's the better model than the closed approach. And then I'm sure you've seen some of the interesting user-facing products that we've launched as well, like Gen AI agents that interact with our users and have personalities and can have conversations. We've created AI-generated sticker apps, so you can create stickers on the fly and WhatsApp's one of my favorite AI tools. I can just tell the tool, Hey, make a sticker that looks like this and boom, I can drop it in a chat thread, it's my favorite thing to be honest. We have assistants out there that sort of operate similarly to ChatGPT and can answer questions and build itineraries and do all these kind of cool things that'll be widely released here real soon.

And we have AI Studio, which is sort of like our homegrown tools for building AI tools that we're also going to make available for people who want to build their own tools and technologies. So lots of product development in this space, not to mention the Metaverse AI-driven engine, which is what it is, and Meta's new Ray-Ban glasses, which I am obsessed with and think they are going to transform people's lives, especially people with disabilities or with special and unique requirements to interact with others. So lots of stuff happening at Meta around AI, really interesting products in the pipeline, and if you just Google Meta and AI, you're going to find lots of cool information and specifics about what they are.

Chris Willis:

So Pedro, let's talk for a second about defining discrimination because we can talk generally about, well, we don't like discrimination, we don't like bias. But before we can understand whether we have a bias or discrimination problem, we have to come up with a measurement for it. And I think there's considerable jeopardy for companies to adopt the wrong definition of bias or discrimination. But really I'm interested in hearing your take on that. So from your standpoint, how would you go about trying to define bias or discrimination to give ourselves a measuring stick against which to evaluate some AI process or model?

Pedro Pavón:

Yeah, look, it's a really hard question, and I think expecting private corporations and non-government organizations to create definitions by which potentially enforcement activities are going to be waged is not a good ask. And I say that in the context of I think there's a responsibility for regulators to tell us what it means, right? And there's lots of examples of this in the law. Employment discrimination is not defined by employers, right? It's defined by governments and government agencies and they tell us what the rules are. So I think there's a huge role for government to play into defining this.

That said, I think there are some really important universal concepts that need to go into the definitions that do get created, and I'll talk about a couple of them here really briefly. I think one of the obvious components to me, and this is Pedro's point of view by the way, not Meta's, let the Meta lawyers give their thoughts on behalf of the company. But for me, one of the obvious and necessary components of any definition of bias and discrimination is addressing the amplification of stereotypes, right? That is a basic fundamental building block of any definition.

Understanding whether or not something is amplifying particularly negative stereotypes that affect marginalized groups is really important as a component to any definition that comes up. The harmful creation and proliferation... Excuse me, the creation and proliferation of harmful content, I think particularly from a discrimination standpoint, that's a really important component to any definition, which is the technology hurting people or creating content that hurts people or hurts people's ability to understand the world in non-harmful ways. That is an important component to any definition we create. And I think the last one is about really representation. Does the technology or does the whatever surface we're applying a definition to create inadequate or representation of marginalized groups or disproportionately negatively impact those marginalized groups? So I think those are the building blocks of any definition that could be created, right? Amplification of stereotypes, the proliferation of harmful content, inadequate representation of marginalized groups, and disproportionate, disparate negative impact on marginalized groups. I hope that's helpful.

Chris Willis:

It is, and I think what you've touched on there, I think is a good opportunity for us to say to the audience that at least as far as I'm concerned, there can't be a single yardstick for measuring bias or discrimination in all use cases because what is a negative stereotype or what is under-representation, et cetera. The things that you just talked about will be different in different contexts. So for example, if I have a credit underwriting model that we would expect a different yardstick for that versus a model that predicts whether someone's interested in dining at a particular restaurant, right? Because dining at a particular restaurant is highly locally based, whereas a credit underwriting model may not be. And so there's endless use cases of AI. And I think the lesson here, and what I was really hoping to and did get from your comments was that we can articulate general principles of how to detect and define discrimination, but there's no one size fits all yardstick for it. I believe that to be the case. I also believe that's why we're not going to get one from the government, by the way.

Pedro Pavón:

Yeah, I think you're probably right on all accounts. I mean, I think your point is really well taken and it's specifically oriented around harm, for me, this idea that one of the ways to avoid harm is to ensure that algorithms are... Their failure rate is extremely statistically low. Seems like a good idea when you first say it, which is algorithms getting it wrong being extremely low is a good bar for rolling out and allowing those algorithms to interact with human beings in whatever context they operate. But I don't think that's true in all contexts, which is sort of like my reaction to your point, which is context really matters when we are talking about harm. If Siri, Siri is Apple's sort of like source, AI agent that everybody's heard of, if Siri 95% of the time puts me at the right address or... That's maybe a bad example, but puts me at the right address for a sushi restaurant, that's pretty good.

And I think that's probably a decent or even 98% of the time that's pretty good, and I think there's very little harm in that 2% for me. There can be some really bad harm, but generally there's very little harm. So let's say 98% is the bar for Siri to get me to the right sushi restaurant, I think that's okay. Now, if there's an algorithm deciding whether people are approved for parole or not or for the death penalty or not, or for one sentence over another or not, the accuracy of that algorithm and the reduction of bias in that algorithm, the stakes there are really high. So 98% success rate is actually not good because 2% of the people whose the algorithm is making decisions about their freedom about are being discriminated against or harmed unnecessarily. That's a really super high level example of how I think just context really matters and how we think about whether or not harm is being done.

Alison Grounds:

Well, and it's interesting, Pedro, I think it's also interesting to think about the... The 2% could be really harmful, but what's the alternative. If those decisions are being made by humans, are we less discriminatory than the algorithm? So it's interesting for me to think about what's the benchmark, what are we comparing it against? Because in some instances, even though there's still an unacceptable rate of harm, it may be less than the alternative. So I think that's an interesting angle to think about. And I'm curious from your industry leader thought perspective, what is the metric? What's the based on, not just the context, but what are you comparing as the alternative?

Pedro Pavón:

I think what you're saying is if human beings get it wrong 90% of the time and the algorithm gets it wrong... Excuse me, get it right 90% of the time and the algorithm gets it right 95% of the time, should we rule it out because that's an improvement. That's not a legal question, right? That is a really ethical, moral question that I think as a society, lots of stakeholders have to participate in the discussion to determine what the right thing to do there is. And I don't think privacy professionals specifically are well suited to answer that. I think it requires a coalition of lots of experts and lots of input from lots of people, especially the people most likely to be harmed by that decision-making process to determine what is the actual appropriate way forward. One of the factors to consider, Alison, is accountability.

So the human bias might be higher, but the ability to hold them accountable might be higher. I don't know the answer, we're speculating here, but it's an angle to think about. What I think you've done though is layered another very important issue onto the one that I was talking about. So I'm talking about whether or not the relationship between harm and accuracy and effectiveness, and then you're layering in the relationship between the algorithms harm and other methods of coming to the conclusions or to the decisions or recommendations that we need to come to for whatever purpose. And so what you're really probing and giving an example of is how complex and how many different issues are involved in deciding what the right path is.

And I think unfortunately, well, fortunately for lawyers and ethicists and people thinking about this every day, there's going to be a lot of hard questions we're going to spend a lot of time trying to figure out, and a lot of people are going to make careers around it. I think unfortunately, mistakes are going to be made and we're going to get it wrong, and we, I mean everybody. And people are going to be harmed, and I think everyone should rally around the idea of reducing that to as low of a probability as possible.

Alison Grounds:

Well, and you made another great point as to the accountability piece, it may be easier to hold humans accountable, I mean, that's debatable too.

Jim Koenig:

Maybe.

Alison Grounds:

But theoretically it may be easier to hold humans accountable for discriminatory or biased decisions as opposed to how do you hold accountable the AI or the algorithm. And I'm curious, just from an industry perspective, what are you seeing as the best practice within the industry and who kind of owns that and who's accountable internally for the results of the use of AI?

Pedro Pavón:

I think the question is inside of a company who's accountable, right? And I think the answer is everybody working on the technology, okay. That's the super elementary school answer, but you're going to have stakeholders that are going to have a lot of active and direct ownership here. Obviously the AI engineers and the data scientists working on the technology itself have accountability, I think, and making sure they understand that the stakes are high for the decisions they make as how they design these technologies and bring them to market, I think is important, right? Companies are now building and creating ethical AI and fairness teams. Salesforce was way out ahead of this and was doing this years ago, as I mentioned, Oracle too. So making sure you have people whose job include responsibilities to think about fairness and ethical use is really important, especially if you're a fan company or OpenAI or one of these companies, sort of like the forefront at the frontier of the deployment of all of these technologies.

I think legal and compliance departments obviously have a role to play, and we don't need to deep dive into that because I'm guessing lots of lawyers are listening here. But ultimately the top executives in the board also have responsibility and it should be held accountable for the decisions the company makes. And I think one of the ways to build that accountability is to have the appropriate internal channels like we have for tax and for privacy and for other areas where there's board reporting and executive level reporting on a routine basis. Building the same for AI is going to be really important. So to make sure that not only the company is holding itself accountable, but if the company or organization makes mistakes, the world can hold it accountable as it should.

Chris Willis:

So I think that discussion is a perfect segway into the next topic that I wanted to talk to you about, to sort of take our discussion in a little bit more practical direction. I think it'd be great for you to share, Pedro with the audience sort of an example from any of the companies that you've worked with, what was a particular scenario where you saw your company take all the right steps to responsibly adopt some AI process or algorithm?

Pedro Pavón:

Yeah, I'm really proud of what Meta has done around Llama 2, right? And I hope most of the folks listening and tuning in have heard of it, but I talked about a little bit earlier and it's sort of our open source, large language model that we've made available for people to innovate on and use. There is a lot of literature that Meta has pulled out about our launch there, but I'm really proud of the work we did to do that launch. I didn't lead that work, but my team did participate in it and I think it had four components I want to flag for you that I think were really important in the success of that launch and hopefully the success of the technology being adopted by lots of developers. One is there was a tremendous amount of red teaming that went on before it was released, right?

And red teaming is anti-talk for human beings fine tuning the model and getting involved in making sure that it's safe and it's being tested adequately before it's launched and after it's launched. I think the red teams at Meta did a lot of work to generate adversarial prompts and to facilitate fine tuning that I know will result in a much safer deployment of Llama 2. I think our transparency framework for launching the model was also exemplary and industry leading, right? There is a lot of technical explanations out there available for Llama 2 that I don't have the know-how or skills to understand, but I think Meta optimized... Excuse me, yeah, optimized for maximum transparency and you can go see the research papers and the disclosures the company has made around not just how great the model is, but a lot of the challenges and issues experienced during the build.

Right now on our website, you can go read it. We also released a responsible use guide. So sure, we made the model available to the world for use, and along with it we provided a guide for how to use it responsibly and how to ensure that you're using it safely. And I think that also is a cutting edge industry leading approach to providing a model publicly for wide use. And I think the final component that I want to highlight is our acceptable use policy. Lots of people spent tremendous amount of time building an acceptable use policy in place that prohibits certain use cases and helps ensure that Llama 2 is being used fairly and responsibly.

If you want to read that acceptable use policy again, all you have to do is go to the Llama 2 website and see it. So those four are the pillars. There's a lot of information in there, but I think it shows an example of a good launch that was done responsibly using the best tools and information available at the time of the launch to ensure that what we put out into the world can be used safely and responsibly by millions of people.

Jim Koenig:

Pedro, that's great. And so maybe a way to think about that as we talk about practices that companies that are listening here that are not an AI first company, they're consumers and users of the technology. What's your thought for them about where to start? Are AI controls a new unique silo or can you leverage privacy controls and procedures for some of the things that are a good place? Again, I mentioned before, leveraging for AI privacy safeguards like a data protection impact assessment, but making it integrated, compliance for privacy, security and AI where you mentioned the acceptable use or I said the ethical data collection, usage sharing charter. It's a high level, where do we start? What lines are we not to cross and what goals are we to strive for and head in that direction or human oversight to prevent automated processing or bias that some of the laws are starting, at least to get started. What do you think about training? What are some of the safeguards that you think that companies should start with and is privacy a good starting place?

Pedro Pavón:

Yeah, let me tackle that second part first. I think we, when I say we now, I mean the collective privacy apparatus. So the privacy lawyers, privacy professionals that work inside companies and organizations have built really great scaffolding for product review as it relates to specific privacy concerns, whatever those are, you mentioned DPIA's which are data protection impact assessments, and I think they provide a really good runway or track for which to layer in AI review, okay?

And so I think the scaffolding is there. I do think the AI reviews that are necessary to ensure the responsible deployment of AI tools and technologies that any level within organizations require a broader coalition of experts than just privacy experts. So what I hope I'll see is not privacy people parading as AI experts and then just incorporating AI checklists into DPIA's, but instead, but privacy experts building out their own expertise around AI certainly, but also inviting AI experts into the deep data protection and impact assessment as we layer in AI components to that process to ensure that the right types of experts are determining whether or not the AI principles or the AI guardrails that each company decides for itself it's going to implement are being met and that the right benchmarks are being hit before products get built, released, or updated out in the world.

To go to the first part of your question, what the specifics of those reviews are, I think lots of companies and organizations are going to differ there because as Alison and I discussed a few minutes ago, all AI's are not the same. And so the standards are going to probably vary even within companies about different AI tools than what's expected of each of them. Are there some minimum basic guidelines? Sure, the White House has put some out. The EU AI regulation is going to create a framework and we should figure out how to build a review model from those, but I think companies and organizations should also make sure that they are customizing whatever review process they layer onto the data privacy review process for AI to ensure that they're not just creating checklists to meet legal compliance, but actually building responsibly in a way that speaks to the technologies that they themselves are building and deploying. So that's really important.

Alison Grounds:

Well, I think that touches on, Pedro, another issue we wanted to address, which was the actual policies for companies and their own use of AI tools. So it's not just is the tool working as intended, but how are we training and overseeing the use of the tools so that they're being used for their intended purpose within a company like Meta certainly. I mean the law firm is a little bit easier, but so what are you seeing in that front in terms of how to ensure the appropriate use of the tools and AI governance policies around using them?

Pedro Pavón:

That's a great question. So as an old contracting lawyer, I remember DPA's, data processing agreements, and these were created and designed as a reaction to regulatory pressure in Europe and in all parts of the world over the last 10 or 15 years to ensure that companies are being transparent about their data protection practices, not just on their own practices, but with their partners and their contractors and whatever. What I saw at Salesforce where I was still much more heavily involved on the procurement of tools and not just the building of tools, but also the buying of tools that included AI was an incorporation of AI guardrails into those tools and a requirement of transparency around AI processing for those tools. The biggest thing that drove that before regulation was actually the protection of trade secrets and having partners use your data to build their own models or optimize their own models.

That's what actually started the incorporation of AI into DPA's, was like you can't use our stuff to train models that you're going to then sell to our competitors to make their business thrive more, right? Now that sounds like a very self-serving intention, and it is in a lot of ways, but what it did do was open the door for AI governance being reflected in commercial contracts between vendors and their customers, right? What we're seeing in the last 12 months, and again I'm 12 to 18 months and I'm less involved than I used to be in this process, but is a explosion of contract language between vendors and their customers around exactly how AI processing works between the companies. And I think that will serve users and consumers and people affected by the B2B use of these tools. In many ways, it'll allow the companies adopting these tools by these vendors to be more transparent with their users about what they're doing.

It will allow for liability frameworks to be well articulated to ensure if something goes wrong, we can figure out who's responsible. And I think it creates a surface for people to object ultimately, right? Ultimately, people will be able to have higher understanding of what's happening in the background with their information, and I think that will drive transparency and control over time. Regulators are probably not going to wait for this to happen organically, and they're going to say, you got to do it. And I think that's fine, but I think that's the trajectory. I hope that answered your question.

Alison Grounds:

Absolutely, yeah.

Jim Koenig:

That's great. I mean, look, I'm not certain law firms businesses, many are going to be building curated databases that they use, their own internal, licensing others using different technologies and tools to layer it in there. Right now, people are just looking at the AI tools that are based on maybe the public or data information and just don't have a vision of that future where you're competing against curated databases and as law firm marketing's going to change. We're not going to send out articles on every single thing under the sun hoping someone notices if they need it for litigation because the second it's out there, everybody else in their model consumed it.

And so how do we re-change the method and approach for engagement and getting people there. Well, enough about law firm marketing, not your challenge. Let me ask about something else of concern, I mentioned them before. There's a lot of interest around deep fakes, authentication and other risks, and so whether it's deep video fakes and audio fakes or other things that perpetuate other scams, what do we do to prevent bad AI? Is it watermarks that are coming? But what are things that maybe we'll see consensus around an implementation to help the bad actors stay away from the progress and innovation that good companies are trying to do responsibly?

Pedro Pavón:

Yeah, it's a great question. Look, I think watermarking has a role to play and it is out at the forefront of the discussion right now, which is helping people understand that the content they're seeing was either created by AI or the creator used AI to make it, right? I think that is important for people to understand what they're looking at, right? Me saying something was made with AI doesn't inherently mean that it's fake, doesn't inherently mean that it's bad, right? And so we have to be really careful as we champion watermarking, and I say we, industry, right? And as governments and regulators also champion the value of watermarking that we don't oversell what it's designed to do, which is basically just telling people that, AI was used, right? I mean, we didn't do this for Adobe Photoshop. I don't have to say I used Adobe Photoshop to augment this photo before I put it on my LinkedIn profile.

I think what's a little different here is just the ease of with which you can augment and the radical nature of some augmenting, and also it's just the scale at which how many people are going to be able to augment. To use Photoshop, you have to know what you're doing to augment my image or any image with AI, I basically just have to type some commands into a tool and it spits it back out. So I think those are things that drive watermarking and it's really important. But if we understand the limitations of watermarking, which is it's not really a protection measure, it's a transparency measure, then you got to look to figure out, well, what are we going to do about all the bad actions that are going to happen? And I think you talked about this already, but developing deep fake detection is going to be really important, right? And how we're going to do that, is we're going to use AI's to combat the AI's, that's the answer, right? We have to figure out how to create tools that can ID deep fakes and flag-

Jim Koenig:

I saw a movie like that.

Pedro Pavón:

Yeah, exactly. And flag them as such. So I think lots of companies are investing really heavily and figuring out how to build technology measures to at a minimum identify deep fakes. Same thing goes for anomalies. And this is not about content, this is about monitoring networks and user behaviors and patterns to find AI threats in your information systems. This is not as sexy as deep fake photos for political news or whatever, but identifying AI's that might not have permission operating in your IT infrastructure is going to be the next battleground for cybersecurity. We're going to have to think about that because the AI tools are going to do a lot of that work too. And the privacy and bias and discrimination implications of having machines fight each other in the background of our daily lives is going to be significant. I think encryption has a role to play here, right?

As AI's get smart and figure out how to unlock access to information systems without permission, making sure that organizations have deployed safeguards to ensure that even if that information is accessed, that it can't be read. It's going to be really important. So more investment is going to have to be made there. And then here's some old classic stuff for you. We're going to have to teach people what's happening and you know how we do the annual privacy-

Jim Koenig:

Training?

Pedro Pavón:

At every company? Yeah, we're going to have to do some training-

Jim Koenig:

Training?

Pedro Pavón:

And just help people understand what the hell is going on.

Jim Koenig:

Wow.

Pedro Pavón:

Yeah. And that's not just employees, that's not just do the employee training. It's teach your users, teach your end of the line stakeholders, your customers, like what the AI-related risks are and how AI fraud is perpetuated, and what other preventative measures they can take as participants in the ecosystem to protect themselves and others. That's going to be really important. And the last thing I'll say again, is also kind of old fashioned, but it still has application, which is incorporating AI into your incident response plans in your organization, figuring out what tools you're going to use and figuring out what your response plans are going to be for AI-enabled cyber attacks and AI-enabled fraud and misrepresentation and deep fake. All of these things are not necessarily related, but they have to all happen in unison for us to build the safest and fairest and most inclusive ecosystem that includes AI's deployed all over the place as I think we're headed for.

Jim Koenig:

So you mentioned compliance oriented versus innovation forward. So let's do a mental exercise question for the last one as instead of talking about laws, regulation, compliance practices, all the things that we thought we came to this call interested in, why don't we end on one and what's the future look like 5, 10, 50 years from now? How's AI helping us and change the world in ways that we might not have thought about yet?

Pedro Pavón:

Well, let me give you a way that I think right now is available and put a plug in for a Meta product. The Meta Ray-Ban glasses, the newest version that came out a few weeks ago. If you put those glasses on, I just came back from Japan, I don't speak Japanese. If I would've already purchased my Meta Ray-Ban glasses, I could've moved through Japan with those glasses on my face, and a little voice would be telling me every sign I looked at a train station, every sign I looked at on a building, every time I had an interaction with written text, it could have translated it for me on the spot. Think about the safety implications of that, for me. Japan is a very safe place, but let's assume I'm in somewhere maybe less safe and I'm about to make a turn into a dangerous alley that has a warning sign on it and I can't read it, right?

Just think about that, think about the convenience implications of that technology. Now imagine that I have a visual impairment. I actually can't see the signs clearly, but the glasses can, and they're walking me through the space that I'm in. The implications of that technology for humanity are tremendous and tremendously positive if we keep building it responsibly and we keep making people feel safe about their use. That is not the future, that is the present, and it can only get better. And I'm really excited about continuing to develop technologies like those that will help people navigate the world more safely and with much more detail. Now, jumping ahead, 50 years, like you said, I think the possibilities are really great. We know the challenges that humanity faces, and some folks maybe don't see them as clearly, but I do, which are climate change is a real thing, at least in my mind.

The possibilities for AI to help us solve the complex problems of keeping the world a livable place are incredible, okay? The possibilities that AI provides us for wildlife conservation and understanding the circumstances necessary for animals to continue to share this world with us are tremendous. We talked about training a little bit earlier, think about what AI and AR, VR augmented training can do just for safety. Like training for spaceflight, using AR and VR, training for medical procedures using AR and VR, training for military operations using AR and VR. Just increase the safety and efficiency of those activities significantly. Imagine you got to teach your fifteen-year-old how to drive, and the first lesson can be in virtual reality. Well, they're not going to crash your Cadillac, so that's at least one good advantage, right? So I mentioned space exploration, I am not an astronaut, but I can see just the automation that AI can bring to tasks and decision-making in such a complex endeavor, how that can improve safety range and capacity.

And then lastly, the one that excites me the most is discovery. There are a lot of really smart human beings that have discovered a lot of really interesting things since the development of, I'm going to use this word boldly, like science, right? Well, imagine what Albert Einstein could have done if he had an AI assistant crunching some numbers for him so he could just stay focused on his imagination and the creation part of his ideation. I am really excited for that little Einstein who's out there in somebody's house right now in elementary school, being able to use AI to just have a more expansive view of the world and be able to think more creatively. And my great hope is that that little Einstein is a she, and that she transforms our world into something much better than we can imagine right now, I hope that's helpful.

Jim Koenig:

That was awesome, Pedro. I don’t know about Einstein, but I knew Edison had Nikola Tesla, so that works out, it works out sometimes. So Pedro, thank you very much for participating today in our discrimination and bias and AI round table and fireside chat. Alison and Chris contributions and your superpowers definitely shine through. Thanks very much, everybody. Take care.

Pedro Pavón:

Thanks for the invite, guys. Really had a good time.

Brett Mason:

Thank you to our listeners for listening to our fireside chat today with my Troutman Pepper colleagues and Pedro Pavón from Meta. I hope that you enjoyed the conversation, that you learned something more about AI bias and discrimination, and that you took away some best practices that businesses can incorporate. Please don't hesitate to reach out to me at BrettMason@Troutman.com with any questions, comments, or topics, suggestions. You can also subscribe and listen to other Troutman Pepper podcasts wherever you get your podcasts, including Apple, Google, and Spotify. We look forward to having you listen to our next episode.

Copyright, Troutman Pepper Hamilton Sanders LLP. These recorded materials are designed for educational purposes only. This podcast is not legal advice and does not create an attorney-client relationship. The views and opinions expressed in this podcast are solely those of the individual participants. Troutman Pepper does not make any representations or warranties, express or implied, regarding the contents of this podcast. Information on previous case results does not guarantee a similar future result. Users of this podcast may save and use the podcast only for personal or other non-commercial, educational purposes. No other use, including, without limitation, reproduction, retransmission or editing of this podcast may be made without the prior written permission of Troutman Pepper. If you have any questions, please contact us at troutman.com.