FEDERAL TRADE COMMISSION

COMPETITION AND CONSUMER PROTECTION

IN THE 21ST CENTURY

Tuesday, November 13, 2018

9:00 a.m.

Howard University School of Law

2900 Van Ness Street, NW

Washington, D.C. 20008

FEDERAL TRADE COMMISSION

I N D E X

WELCOME AND INTRODUCTORY REMARKS

MR. GAVIL: Good morning, everyone. My name is Andy Gavil, and I’m a Professor here at the Howard University School of Law. On behalf of Dean Danielle Holley-Walker, my faculty colleagues, and our students, I’d like to welcome the FTC and all of you to Howard for Hearing Number 7 of the FTC's hearings on Competition and Consumer Protection in the 21st Century. We are very happy to cosponsor today's event, and I want to thank the FTC and the many people at the agency and here at Howard who have worked hard over the past few months to organize these hearings.

As you all know, today's topic is Algorithms, Artificial Intelligence, and Predictive Analytics. As is immediately evident from both the list of questions the FTC has posed and the agenda for today and tomorrow's programs, these hearings have been purposefully designed to take a broader and more interdisciplinary perspective than any of the previous ones.

Moving well beyond the usual collection of academic and practicing economists and lawyers, FTC staff have assembled an impressive collection of academics, public servants, technologists, scientists, engineers, and industry leaders, but of course, there's still lots of lawyers and economists.

The goal is to educate the agencies and the broader competition and consumer protection policy community so we can all obtain a better understanding of the technologies that are transforming our economy, as well as our political and social environs. We'll hopefully learn more so we can better understand the business models and practices of our time and so we can differentiate myth from reality, promise from near and long-term prospect.

The ability to take on this kind of prospective study is a hallmark of the FTC and one of its great institutional strengths. It is especially fitting that such a forward-looking approach is being taken here at Howard. Only two years after Howard University was chartered by Congress in 1867, this law school was founded with the aspiration of producing lawyers who would lead the future fight to realize the full promise of the reconstruction amendments to the Constitution of the United States.

Next year, we will celebrate our sesquicentennial, and for that occasion, instead of looking backward, we have selected a theme that looks forward, “The Next 150." As is true for the FTC and for today's hearings, any institution that fails to look forward is bound to fall backward.

In closing, please note that the event is being photographed and webcast and will be posted on the FTC's website, and that by participating all attendees consent to those conditions.

Please also note that our students will be coming and going throughout the day and are available to answer your questions. Please get to know them while you are here and feel free to seek them out if you have any questions or concerns.

Finally, it's my great pleasure to introduce our first presenter. Our scheduled presenter, Michael Kearns, has been slightly delayed, so we're going to start with John Dickerson from the University of Maryland, and hopefully Michael will arrive in time to follow John. Again, welcome, thank you, and enjoy the hearings.

PRESENTATION: INTRODUCTION TO ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS

DR. GOLDMAN: Hi, I’m Karen Goldman. I'm an attorney in the Office of Policy Planning at the Federal Trade Commission, and I just want to introduce you to John Dickerson, who is an Assistant Professor in the Department of Computer Science at the University of Maryland, College Park. Welcome.

DR. DICKERSON: Thank you, Karen. It’s a pleasure to be here. I am John Dickerson, I'm a, I guess, third-year Assistant Professor at the University of Maryland and right up the street in College Park, and today I'll be talking about an introduction briefly introducing the audience to algorithms, AI, and predictive analytics.

And so for this talk, I'd like to start with a motivational quote which sounds like it was written a long time ago, and that's because it was. So “although machines can perform certain things as well or perhaps better than any of us can, they infallibly fall short in others... by which means we may deduce that they did not act from knowledge, but only from the disposition of their organs.”

And this sounds old because it was written a long time ago. It was written by Descartes, who was a philosopher and mathematician in the 1600s. So quite a long time ago, folks were already thinking about what does it mean to think, can we mechanize thought?

Another famous philosopher from the 1600s, Hobbes, states, “Reasoning is nothing but reckoning.” So reckoning here is just a reference to mathematics. So reasoning is nothing but mathematics essentially.

And so some time passed, 1600s, 1700s, 1800s, until the 1900s, when some breakthroughs occurred in logic and mathematics and philosophy. Folks like Boole, folks like Hilbert, made some breakthroughs in the formalizations of mathematical reasoning. So recall, we think reasoning is nothing but reckoning, and now we can reckon perhaps with mathematics.

So there were some proofs showing that some hard limits -- there are some hard limits to what mathematical reasoning can do, but subject to those limits, folks like Alan Turing came around, Church came around and said there are certain machines -- simple machines -- that for any of these mathematical reasoning problems, subject to these limits, we can create a machine that can do this.

So this is nice, this builds on now hundreds of years of philosophy and mathematics, but the general pitch here is that if intelligence can be simulated by mathematical reasoning, that is reasoning is just reckoning, and mathematical reasoning can be simulated by a machine, then can a machine simulate intelligence?

So AI, artificial intelligence, the word was coined by John McCarthy in either 1955 or 1956, depending on how you count, it’s '55 in a proposal, to fund the Dartmouth Summer Research Project on Artificial Intelligence. And you'll hear this called the Dartmouth Conference. This occurred in the summer of 1956.

And there are some fun quotes in there saying basically we can solve artificial intelligence in three months or we can solve artificial intelligence in one generation, but the one I'd like to pull out is that every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it. So even in the 1950s, 1960s, folks were making statements like this.

So a quick spoiler, this hasn't happened yet. We can just shut this down right now. But, progress has been made. So how does that progress occur? Well, this is a cycle of basically R&D progress that you'll see repeating in the AI world, and this has happened since basically 1956, where some new advance, maybe a new technique, new hardware happens. Fast progress is then made on old, hard problems. So it could be a new mathematical technique, it could be new hardware, GPUs, these graphics processing units, are one of the main drivers in the current sort of fast progress being made on problems that we're seeing now.

But eventually you start to hit road blocks. And at this point, the academic community, the industrial community starts to get pessimistic, this bleeds into the press, and at that point, everyone is pessimistic about progress, funding dries up, progress dries up and so on. We wait until the next large advance.

And so this is the cycle that occurs in most sorts of verticals. It occurs in AI research as well. In AI, though, we call it a cycle of basically AI summers and AI winters. The winters are when funding dries up and nothing happens; the summers are basically what we’re going through right now, where we’re seeing large advances driven by sort of recent hardware and mathematical advances.

So this is a bit pessimistic, this cycle, but like I said, progress has been made. So this has been cycling for arguably maybe six or seven times since the 1950s, but every time we go through this loop, progress is made, new problems are solved, and new problems are encountered.

So what is AI? AI, many definitions, the one I'll use here is the ability to process and act based on information via automation. So we can break this down roughly into four segments. One is perception. I want to be able perceive the world around me. That could be the physical world; that could be the virtual world. I want to be able to learn something about it. So I get some signals about the world, then I learn something about them. Maybe I learn a model.

I want to abstract and generalize that model so that I can use it in other situations. And what do I mean by use? Well, maybe I can reason about this information, I can reason using my model and then act within the world. Again, that could be virtual, that could be physical.

So if I can create this automated system, roughly, I have created what we would call AI. So let's keep moving through this history of AI until we are where we are today. Roughly we can split AI research into some first-wave AI, second-wave AI, and then maybe 2.5 or third-wave, which is where we are right now.

In this first wave, primarily, researchers focused on what is called search. So this is either searching through a potential solution space, some quick examples, chess is a good example here where we had, say, Deep Blue beating Kasparov via a sophisticated algorithm that did search through using domain-specific heuristics, expert knowledge, for instance. Folks who played a lot of chess encoded heuristics into the search algorithm; it would search through the solution space to find, say, the next move to play.

Now, another hallmark of first-wave AI is something called expert systems. And this also relies on basically bringing in a lot of expert knowledge to form some sort of large database of rules, of knowledge, of facts about the world, using some sort of inference engine, typically based on logical reasoning, to make new sort of conclusions based on these facts, and then some sort of action, I/O system to interact with the human. So this is basically the world up until maybe the ‘80s in terms of AI.

Now, there were some large successes here, so one example that I used earlier is this chess champion falling to basically a sophisticated search algorithm. And there are many more. And, in fact, techniques from first-wave AI are still used in practice, but they're decidedly brittle and they really don't have any real learning capability. So they're really sort of a function of just the knowledge that you encode into them.

There's a huge overhead to encoding that knowledge. Right, I have to ask, say, every member of the audience and everyone watching to tell me all the facts that they know about the world and then I have to store that somehow, and that might be brittle and that might not be generalizable. They're very, very brittle systems, but they do allow me to do in-depth specific reasoning. Right, if I ask a bunch of experts for facts on a specific vertical, then I can do a lot of fast automated reasoning about just that vertical. So that can be good, but it's very difficult to generalize.

And if you recall back to that earlier slide, we want generalizability, we want abstraction because we want to create some system that's able to encounter new environments and still act in a reasonable way.

So in terms of those four boxes, first-wave AI generally does perception reasonably well in the sense that I have asked all audience members to give me facts, and it can do reasoning and acting reasonably well, but it won't learn and it won’t generalize.

Now, there were some transition points in multiple areas of sort of AI research. One of these is something called natural language processing, which says, can I get a computer to ingest, say, raw text or can I get Alexa to ingest signal from your voice and then have it understand that in some sense. So in natural language processing, up until about the late 1980s, most of the rules for doing this sort of translation or understanding were done via hardcoded sort of expert rules.

Around the late ‘80s, probabilistic models started to come into play. Okay, so this is going to sound more like machine learning like folks have maybe heard about in the press. These are models that ingest, in this case, large text corpora and learn patterns in that data.

To look at a different vertical in AI, so autonomous vehicles rely heavily on something called computational vision, which says, hey, I have a video image, can I understand what’s going on in that image. Say I'm a car and I’m driving along, and I have a still image of the road in front of me, can I understand that there's a stop sign and a pedestrian and dog in front of me and so on. So in autonomous vehicles, in the mid-2000s, DARPA ran what they call a Grand Challenge, in fact their first Grand Challenge, which asks, can I create a vehicle that can drive some hundred-plus miles across the desert autonomously?

In 2004, no vehicles completed this task. In fact, I think the longest trip that a vehicle took was something like ten miles. And these vehicles relied heavily on hand-coded rules that say something like, in general, when you're, you know, ten degrees away from the sun and you're driving forward at a particular speed, then a shadow is going to be a shadow instead of a rock with some set of features associated with it. And, again, this is a very brittle system. This is not going to generalize very well.

But then in 2005, five teams completed the entire trip, so 100-plus miles. And this is because they started using these probabilistic models. And, in fact, you can see the general manager for the program, Strat at the time, had a fun quote: “[Vehicles] were scared of their own shadows, hallucinating obstacles when they weren't there.” And this is for those prior systems. And then probabilistic models allowed them to get around this.

So you can see similar transition points throughout all core AI areas, in the late ‘80s, in the ‘90s, up and through basically the mid-2000s. And this happened because of three things. One is computational power increased, and this is the story of basically computation since the ‘40s or ‘50s. This has played a driving role in AI development as well.

Number two, storage costs decreased. I don't have to pay a lot of money to store a lot of data. And, three, everyone in this world now relies on statistical models, maybe with some expert input, but still statistical models.

So this takes us into the second wave of AI, and there's no hard date for this because it happened differently in different verticals in this world. Here, we're relying on this assumption now that we've learned the hard way, multiple times, that encoding all knowledge explicitly does not work. It doesn't scale. It's very brittle and it's very difficult to handle uncertainty.

The new idea is that we should create a general statistical model for a problem domain. We should create a statistical model for natural language or for a type of natural language or for autonomous driving, a type of autonomous driving. What do we do with that model? Well, we feed in data from the real world or maybe simulated data until it looks right. And this is going to be characterized by statistical learning.

So the reason why these models have taken off is because if we input a different data set or, say, set of data sets into these models, we'll learn a different model and then we can deploy that in a different environment. So it's much more generalizable.

Now, some examples. In machine translation, for instance, going back to this natural language that we discussed earlier, we can feed in multilingual text corpora to learn relationships between languages. So say we want to translate French to English, one of the early multilingual text corpora came from Canada, where there are rules stating that, say, any government ruling has to appear both in English and

French. And so now we have a mapping between English and French documents, we can feed that into a model and we can learn a way to translate between the two systems.

Autonomous vehicles. We can feed in videos and tests of successful driving into a model and then learn what scenarios are safe or not safe or maybe put some error bars around what scenarios are safe in general.

Face detection, face recognition. I can feed in many labeled faces of people. Here is where the face is, or here is where the face is and an idea associated with that, to learn what a face looks like or to learn what, say, your face looks like.

So these types of models are very good at perception, and they're very good at learning. Remember, we're training these models, these general models, based on a data set, and if we feed in a different data set, we're going to get a different result, so they’re reasonably good at abstraction and generalization as well, so long as your model is general enough and so long as you have enough data. But there is no reasoning or acting. I've made no statements about, say, when one should turn the car in -- turn the wheel in the autonomous vehicle.

So a quick example model. Remember, these are systems that rely on statistical learning to train probabilistic models that will tell us something about the world. A quick example is a neural network. So these appear a lot in the news now, which is why I've chosen them, but they're not a new idea. Indeed, that 1955 proposal where McCarthy defined AI, used the term AI for the first time, also discusses neural networks. I believe they were called neuron networks at the time. So this is not a new idea.

The general idea of neural networks is that one should pass information into this input layer, which you see on the left side of the screen. So that information could be pixels of an image. That information could be something with audio signal. It will cascade through the network, along basically a series of pipes that go through nodes, and these pipes have, say, different widths that can be controlled by a learning algorithm.

And then the final layer of this network that has information flowing through it will create some sort of guess. In the case of, say, classifying images, here we have cats and dogs, it's going to create, say, a probabilistic model of whether or not an image is a cat or a dog. And that gives you some signal as to how good or bad your statistical model -- in this case a neural network -- is acting.

A very general model, so long as we can feed information into it via that input layer and so long as we can judge the output and so long as we can actually learn, so make the network better, using sophisticated optimization techniques, we can use this for many problems and, indeed, that is what we've seen, so long as we can, again, train these models through repetitive sort of optimization algorithms.

So another sort of buzzword that one sees in the press a lot is a deep neural network. Again, not a new idea. These existed, I think, since the 1980s, and they're just these neural networks that we had on the last slide but with more, quote, unquote, hidden layers. These are the layers in between that input and that output. So I can add more and more of these. I can create more piping -- intricate piping between these different nodes to learn new patterns in the data.

And sometimes, indeed, we can stack many, many, many, many, many more nodes, so order of hundreds of thousands, millions, et cetera. So these are very large models. And, again, this is because we have increased computational power and cheap storage.

That idea for deep networks has existed since the ‘80s, but we've seen them taking off in the last five to ten years because of advances in hardware, because of a huge increase in the amount of data that exists. So we have large firms collecting data; we have the government collecting data; and we can now store it cheaply, access it quickly, and because, indeed, from the R&D community, there have been much better methods developed for learning basically how to make a good one of these.

They're hugely successful. They're good at detecting anomalies in data, for instance, credit card fraud. They're good at voice recognition. You've seen Alexa, Siri, Google Assistant, et cetera. They're great at machine translation, language generation, game playing. Some recent high-profile success stories such as AlphaGo playing basically expert-level, Go, DeepStack Plane, expert-level Heads-Up Poker.

Self-driving cars are starting to take off. Video search, audio search, finance, et cetera. These are all success stories in part due to deep learning. Not a new idea, driven by advances in hardware and training them.

Nobody understands why they work very well, and this is a common story in AI as well and this is something that we're seeing more and more appearing, which is humans have sat down, they've designed the network structure, they've designed what those nodes and what those connections between the nodes look like. Maybe they're encoding some domain expertise. There are some known heuristics that you can rely on. There's a trial-and-error process, and maybe actually other AI is actually coming in and trying to train these models or structure these models in a better way, but nobody knows when or why they don't work, in general.

So they work well in expectation, which is why we see machine translation systems, which is why we see Alexa and Siri in households now, but when they fail, it can be very confusing, it can be reasonably catastrophic, and it can be very hard to explain.

And some recent research pushes funded by the DOD, funded by industry, funded by nonprofits, have started noticing that, hey, an adversary can exploit this kind of behavior. When I have a system I trust most of the time but it can be exploited in very odd ways and I don't understand why or when that happens, then I can wreak some havoc in these systems.

So I'd like to take a step back. So now

we've talked about deep learning, we’ve talked about machine learning, and we’ve talked about AI. And, roughly, AI is this sort of four-pillar approach to perceiving the world, learning about it, building an abstract and general model, and then using that to act and reason. Machine learning is just one way to build these models, where we do not focus on acting and reasoning but we focus on perception, on learning, on abstraction, and on generalization. And deep learning is just a specific form of basically representational learning, so it's a type of machine learning.

Right, so every time you hear deep learning in the news, you can replace it with machine learning mentally. It's just a way to solve a machine learning problem.

So some present-day movements in AI, understanding bias and methods for debiasing. You'll hear about this I think throughout today and tomorrow, many of the topics on this slide. So this is sort of a teaser. Understanding bias and methods for debiasing. So if I feed skewed training data into these systems -- remember, these are statistical models that are trained on data from somewhere in the world. If I feed skewed data into the system, then

I'm going to learn something that represents that skewed data. So how do we understand when that happens and can we create systems that still feed in this biased data which might be the only data that exists but spits out a model that is debiased?

As mentioned before, adversarial reasoning in multi-agent systems, learning to act with cooperative actors, learning to act with adversarial actors, so bringing in older fields such as game theory into these new methods for solving those problems. How do I say design -- well, I'll talk about this in a few slides, but how do I design policies as a firm to compete with other, say, firms that are both cooperative and adversarial? Can I do this based on machine learning?

Also mentioned on the previous slide, robustness to noise, robustness to adversarial attacks, both in terms of theoretical robustness and empirical robustness. How do I design automated systems that fail less, that are robust to attacks and that fail more predictably, because obviously these systems will always fail at some point.

And in that vein, explainable AI, there's a lot of money going into this as well because it's very difficult to interpret the results that come out of these systems from time to time, so can we produce human-understandable models that also work well?

And one final move in the AI community has been reinforcement learning. It's a type of machine learning, but it's a type of machine learning that also focuses on learning to act and reason. So now we're getting closer to that initial definition of artificial intelligence. Here we have an agent, maybe physical, maybe virtual, that's going to act within an environment. It's going to receive a reward signal and then maximize total reward. It wants to find the actions to take for any state in the world such that when it takes that action, it is treated well in the future, it receives reward and expectation in the future. And I'll give you some examples of this at the end of the talk.

So here again, again, reinforcement learning, not a new idea, but deep networks have been used extensively here to revolutionize their use and practice. So here we have deep networks that are used to, say, reduce the complexity of representing the environment. Remember, I can't actually write everything down, I don't want to represent every single fact in my computer, so now I'm going to learn some abstraction of the world and then act on that.

So reinforcement learning is taking us closer to what we want to call AI. We have perception, we have learning. These are just like machine learning. We have abstraction and generalization, again, moving toward that. Again, if we train these models on different data, we get a different trained model, and we're starting to move toward reasoning and acting here.

So in the context of this audience, I thought I would do maybe a quick deep dive into a few uses of AI, particularly in something called market design. So markets provide agents the opportunity to gain from trade. Many markets require structure to operate efficiently. Market design is going to tackle this problem via what's called economic engineering. So I put on my economist hat and I put on my engineer’s hat and I put on my mathematician's hat. I’m wearing three hats at this point, but I can use these hats to design a market, how do I structure the market, how do I constrain the market such that I achieve some sort of efficiency goals.

AI is increasingly helping with the design of these markets. For instance, these automated methods can use data to help designers characterize families of market structures. They can be used obviously for predictive methods that anticipate, say, future supply and demand in electricity markets or finance markets.

One example, as a computer scientist, this is close to my heart because a lot of the money in our world comes from this, is using AI in online advertising. So online advertising markets generally match advertisers with consumers. Many billions of dollars, and this is an increasing market, many, many billions of dollars are being used here, and it's a driving force in the technology sector.

Machine learning models in this space right now are being used to divide customers into very fine-grained and automatically generated segments. So no longer just male/female but something far, far more fine-grained than that. That’s learned automatically. They're being used to set reserve prices and auctions based on user modeling and bidder behavior, again automatically.

They're being used to automatically generate the creatives, that is, the artwork that you see pop up on your screen, to automatically generate those, say without human input, to fit a specific customer’s predicted wants. All automated.

Reinforcement-learning-based tools --

remember, this is that form of machine learning that also focuses a bit on acting -- are being used to help advertisers, for instance, bid automatically on these very fine-grained segments. Remember, now we have, say, millions of segments. How do I bid on that, I can use a machine-learning-based model to do this.

Another example, AI in electricity markets. Here, matching supply and demand is extremely important. It relies heavily on demand forecasting. Machine-learning-based techniques are going to provide very accurate demand forecasting, which leads to very stable market prices and more efficient power usage.

Reinforcement-learning-based techniques -- remember ML plus some form of acting -- are going to allow us to activate or deactivate expensive heterogeneous power sources to maintain that stability. So I can predict better demand, I can predict better demand at particular time points further into the future, and then I can make a plan to boot up or boot down particular power sources such that I maintain market stability, such that I reduce brownouts and so on. Again, automated.

And my final example is AI and kidney allocation. This is close to my heart. I've done a lot of work in this space. So here, kidney exchanges, for instance, are an organized market where patients with end-stage renal disease enter and are able to swap donors -- willing living donors -- to receive new organs.

It's a really interesting paradigm that's been around for, say, 15 years now, and it accounts for something like 10, 11, 12, 13 percent now of all U.S. living donations of kidneys. Hundreds of transplant centers are involved in this organized market, in fact, multiple organized markets.

And, here, AI-based tools are also operating. Now, this isn't fully automated, but they are, for instance, semiautomatically and optimally subject to human value judgments, matching donors to patients, both in the U.S. and also worldwide. Here, I've called out the United Kingdom and the Netherlands, but in many countries. They're providing sensitivity analysis at a level that humans cannot for new policies. And they're learning from data the quality of, say, potential matches in this market.

Now, let's return to some open questions and some recent pushes which will, I guess, trigger good discussion for the rest of today and tomorrow. So one is, how and why does deep learning work? So

I've mentioned not a new idea. Neural networks existed since the ‘50s; deep learning existed since the ‘80s. Now we have new hardware and now we have new training techniques, these tend to work very well in expectation, but when they fail, they fail confusingly. Why do they work?

How can we handle incentives of competing agents? All those three market examples that I showed you, firms are obviously going to compete against each other in this space. The government, regulatory agencies have their own incentives as well. Individual participants have their own incentives. How do we handle this, how do we encode other aspects such as fairness, accountability, and transparency into these systems?

How do I ensure that my automated system doesn't marginalize, say, an already marginalized class in the ever sort of increasing hunt for efficiency? How do I even define this? How do I define fairness? This is a classic question in economics that computer scientists are now starting to struggle with as well. How do I implement this in a scalable way, in an understandable way?

Ethical AI, this will be talked about, I believe, later, by folks like Henry Kautz, how do I divide the labor between policymakers, such as those in this audience, who are ethically trained and ethically minded, and technically trained, perhaps ethically neutral, AI and machine learning experts? So I can implement, say, a very sophisticated system, but I need to then produce some sort of aggregate output that I can pass back to policymakers to ensure that this is reflecting the aggregate human value judgments of those who control the systems. How do I do that? And there are close ties in this sort of exploration to the world of privacy and the world of social norms.

So in general, our end goal is to create these systems that perceive the world, learn from it, create some sort of generalizable model and then inevitably learn to act using that model. We're not quite there yet, but there's a lot of hope in this space. But, I'm going to say that maybe this isn't even the biggest problem. The biggest problem is going to be the interplay between these systems and society, ethical issues, societal norms, human value judgments. How do we play between, say, these sort of sophisticated machine-learning-based approaches to what I've shown here on this slide and the rest of the real world? So I'll leave it at that.

DR. GOLDMAN: Thank you very much, Professor Dickerson, for that excellent introduction to the field and for the questions that will be coming throughout this hearing.

(Applause.)

OPENING ADDRESS

DR. GOLDMAN: As I mentioned, I'm Karen Goldman. I’m with the Office of Policy Planning at the FTC, and now I'd like to introduce our next speaker, Michael Kearns, who is a Professor at the Department of Computer and Information Science at the University of Pennsylvania. Welcome, and we're looking forward to your address.

DR. KEARNS: Okay, thank you. So not only am I late, but I also missed the deadline to give slides last week. So there's two strikes against me already. But hopefully the time I would have spent hacking PowerPoint I put into thinking instead, and so I'm just going to speak informally from notes.

So as she said, I'm on the computer science faculty at the University of Pennsylvania. My main research area is machine learning and related areas. I've been in this area since I was a doctoral student in the 1980s, before machine learning was a thing in society. And so I was just asked to give some introductory framing remarks based on the agenda that I saw, which contains, like, lots of topics that are near and dear to my technical and related interests.

In particular, following on the last speaker, in the last few years, I've been thinking quite a bit about ethical and social issues in the use of machine learning and algorithmic decision-making more generally. And I also saw that there are some discussion or a panel about sort of competition and marketplace questions introduced by machine learning. I hope to make some less technical remarks about that because I think that's less scientific to say there but a lot of interesting things to discuss, and also relatedly topics related to consumer protection and abuses by machine learning and AI.

And so what I just want to do with my time is make some informal remarks, provide some personal opinions on these topics based on my own experiences and research, and, you know, hopefully cue things up for the next couple of days for the rest of the speakers.

So as the last speaker mentioned, there has been a lot of discussion really not first in the technical community but first in the mainstream media and society at large, about the many things that can go wrong when applying machine learning and AI and related methods to algorithmic decision-making.

And before I describe -- say a little bit more about what can go wrong and what we might hope to do about it scientifically, I thought I would start by sort of just framing how things can go wrong in the first place. And so one thing you might wonder is, you know, if there is a lending model or a consumer credit-scoring model that exhibits racial bias, for instance, or there's some data analysis or machine learning methodology that leaks personal, private consumer data, you might -- it's a reasonable thing to wonder whether this happens through active malfeasance. You know, are there evil programmers at tech companies who, you know, put in a line in their code that says if the person's race is this then do this; if it's some other race, then do something else; or whether they program back doors into their code that permit privacy leaks.

And there's good news and bad news here. My strong belief, and I think those people who work in the field would say that, no, there is absolutely no such malfeasance going on by evil programmers at technology and other companies. So that's the good news. The bad news is that the truth might actually be a little bit worse, which is these sort of collateral damage or consequences are actually the natural byproduct of applying the formal, fundamental, scientific principles of machine learning and AI. And I’ll say a little bit more about that.

The vast majority of what I think we would call algorithmic decision-making is actually a little bit more specifically almost always driven by machine learning these days. So, in particular, when you think about the algorithms that make things like lending decisions or decide what ads to show you on Facebook or Google, these generally are not what you should think of as hand-coded or programmed algorithms, but rather they're the result of taking data, you know, historical data, whatever that might mean in a given domain, giving it to an algorithm, and that algorithm, of course, trains a model on the data. And then at the end of the day, it’s the model that’s actually making the decisions. It’s the model that’s actually deployed in the field.

And, typically, the algorithm that transforms the data into a model is actually tremendously simple and very principled from a scientific standpoint. So if I had slides, one thing I like to do in forums like this is put up the Wikipedia pseudocode for the so-called back propagation algorithm for neural networks which the previous speaker mentioned. And that pseudocode is literally a simple loop with about ten lines of code in it.

And a real working version of it wouldn't be that much more complicated. And it's doing the most obvious thing you can possibly imagine, which is essentially going through the training data and adjusting the parameters or nods of the model in order to minimize some -- you know, usually accuracy or error-based cost function, okay? So that algorithm is not opaque at all. It's entirely transparent.

Sometimes, you know, when I talk to people who aren’t in the field, they naturally assume that machine learning algorithms -- you know, the code for them might look like something like I imagine the code to a video game like Grand Theft Auto looking, you know, hundreds of thousands of lines of spaghetti code with all these special cases and details, and it's not like that.

So, then, the natural question to ask next is if the complexity doesn't lie in the algorithms themselves then where does the complexity creep in? And, of course, it's from the interaction of the data being processed to produce a model mediated by these very, very simple algorithms, okay? And so the problems arise these days not so much from the algorithms themselves, which, again, are very simple and operating on very basic, kind of well-motivated scientific principles, the problem is really when you work in extremely large complicated model spaces, of which, you know, deep learning is just one and perhaps the most recent example, the sort of space of models has a lot of sharp corners in it, as I might put it, which allow to you kind of optimize the thing that you're trying to optimize like minimizing the error on the data at the expense of other things that you didn't explicitly ask for like fairness or privacy.

And I think if there’s one kind of important adage to understand about machine learning, it’s that basically modern machine learning will not give you for free anything that you don't ask for and specify. And in general, you shouldn't expect it to avoid things that you don't want that you didn't tell it you didn't want. Okay?

And this is, I think, the source of a lot of the kind of violations of social norms and values that we’ve seen by machine learning and AI in recent years. So that's a little bit about what can go wrong. Now let me talk a little bit about -- sorry, that's a little bit about how things can go wrong.

And, so, with that background, I want to talk about, well, what are the different things that can go wrong, and, most importantly, what can we do about them from a kind of scientific standpoint. So, you know, the things that can go wrong are things that I’ve mentioned already, which is violations of things like privacy or fairness or interpretability and transparency, or even safety or morality, if you like. You know, the sort of logical extreme of this for those of you who’ve heard of it is, you know, this sort of parlor game or science fiction thought experiment known as the singularity on which AI, you know, sort of -- AI achieves superhuman intelligence to the point that, you know, for lack of a better term, the robots become our overlords.

While that's a fun thing to think about, I don't know many sane people in the machine learning community who actually think that that's our sort of gravest technological risk anytime soon. All you need to do is come and see what AI and machine learning can actually achieve right now and compare it to humans or other biological species and you will be deeply underwhelmed by what we can accomplish so far. But violations of social norms are, like, already with us now today and on a very large scale, whether we are -- whether we know it or not or whether we’re measuring them properly or not.

And, you know, I think it's important to say to this audience that I think I and many of my colleagues, you know, we do believe that better laws and better regulations are possible and should be developed. And I'm sure that that’s being worked on and is a necessary activity. But I think my opinion is that that will be woefully inefficient in the algorithmic era to actually keep up with the types of violations of social values that we’re seeing because it just -- you know, basically human organizations don't scale, and you can't sort of expect to police the sort of violations I’m talking about with sort of regulatory agencies that are pouring over the decisions made by algorithms on a sort of a human time scale and hope to keep up.

So an alternative approach, which I'm a great advocate of and as are a growing number of people who do technical work in these areas is to design better-behaved algorithms in the first place and to actually endogenize various notions of social norms inside of our algorithms and asking that our algorithms -- that the actual code obey some definition of privacy, some definition of fairness, some definition of morality, if you like.

And, of course, this leads immediately to two very difficult questions. The first difficult question is, you know, how do you define these things as the last speaker said. How do you define algorithmic fairness, how do you define algorithmic privacy? And, then, if and when you can come up with such a definition, it’s going to come at some cost, right.

So if I have some notion of fairness in models that are used to provide criminal sentencing guidelines, my asking for fairness from that model by gender or race will come at a cost of accuracy. What I’m saying is like a tautology. If I sort of -- if I ask myself to find the model in some space of models which minimizes the error period, and then I ask to find the model that minimizes the error subject to your favorite definition of fairness, the error can only get worse.

And so in a model like -- let’s say in a setting like criminal sentencing, this means that a cost to accuracy might mean sort of, you know, hard things to think about. It might mean incarcerating more innocent people, or it might mean letting more guilty people go free. So there will be societal and technical costs to imposing these sorts of constraints on our algorithms, but I think my view and the view of many people in the field is that we have to go down the road, we have to decide algorithms that incorporate these values, we have to talk about what the possible definitions for these values are, and we need to study these tradeoffs between the thing that people optimize for in machine learning, which is accuracy, and the tradeoffs to different social norms.

Okay. And so what I want to do with most of my remaining time is just tell you a little bit about the sort of very active research that's going on in the computer science and machine learning and related communities on this scientific agenda, sort of picking definitions for different social values or norms and actually encoding those norms inside of our algorithms and then trying to study what the tradeoffs will be with, you know, things like accuracy and other more standard objectives.

So let me first talk about the work that goes on in the area of privacy in machine learning, and not just in machine learning but more generally in kind of data analysis and data science. And I think it's helpful to say just a little bit about the distinction between what I’m thinking of as privacy and a closely related and complementary area, which is that of security and cryptography.

So security and cryptography, to a first approximation, is a technology about keeping data locked down. It’s about controlling access to data and making sure that people who shouldn't have access to data don't get that access by basically hacking into a system that they shouldn't hack into. And this is largely the domain of security and cryptography, and that’s one notion of privacy. That’s sort of control of your data and making sure it doesn’t get -- you know, it doesn’t get accessed or stolen by people who shouldn't.

Here, I’m talking about something a little bit different and more nuanced but in many ways is equally as pervasive and important as notions of security, which is the fact that, you know, you have all of this data that’s being collected by different companies and agencies and other organizations. And you might worry about what -- not just sort of, you know, how -- who’s accessing that data but what can be inferred about you from that data that isn’t directly in the data itself.

So the kind of thing that I'm concerned about here is that if your medical record is used as part of a study to build a predictive model, let’s say, for some disease based on symptoms, and then that model is published or used in the field, could it be that the use of that model or the publication of that model, perhaps combined with other publicly available data sets, actually reveals a great deal about your own personal medical status and record. Okay?

And, you know, if you go down the road of thinking about possible technical definitions of this type of privacy, I believe that most of you would eventually come to two kind of, I think, important conclusions or desiderata from any sort of privacy definition for machine learning or data science.

One is that, you know, you need to account for the fact that any particular data set that you want to, you know, keep private in some technical sense, will not be the only data set in the world. And, in particular, that data set might be combined with other data sets that you don't know about or didn’t foresee or don’t even exist yet but might exist in the future.

And one consequence of this that I will

state without proof is that this means that any definition of privacy that it involves anonymization is essentially a flawed definition of privacy, right, because anonymization refers to taking the data set that’s in front of you and doing things like eradicating personally identifiable information.

But the literature and news is, you know, rife with examples where you anonymized one data set, somebody else anonymized a second data set. Those two data sets were combined and then maybe combined with some publicly available information, and your specific data could be backed out of that. You could be, as we like to say, reidentified, or the data set could be, you know, deanonymized as they say.

And, you know, I think many people feel strongly enough about this assertion that there is sort of a saying in the field, which is, you know, anonymized data isn’t, meaning that, you know, whatever you think you did to deidentify individual identity in a data set, that can often be undone through the unforeseen combination of that data set with other data sets.

The other, I think, sort of axiom for any definition of privacy that’s important is that in order to have a definition of privacy that still allows to us do anything useful with data, it's important to isolate, you know, the potential harm that comes to somebody as the result of use of their data in some analysis or model-building exercise versus the harm that might come to them just because data analysis reveals some facts about the world.

So, for instance, if you were a smoker in the early 1950s before there was discovered a link between smoking and lung cancer, well, when somebody did data analysis and discovered that there was a strong correlation between lung cancer and smoking and you were a smoker, that fact does harm to you, but it doesn't matter whether your data was used in that analysis or not, right?

Researchers were going to discover this fact whether your particular data was used or not, and a harm was done to you by the fact that suddenly it's revealed that smoking is bad for your health and you were a smoker and everybody knows it. But that harm was not done to you specifically as a result of the data analysis using your data or not. You were basically -- you know, this harm was going to be done to you whether your particular medical record went into those studies or not.

And so there is a very rich definition of data privacy known as differential privacy that was introduced about 15 years ago and has received a great deal of scientific attention in the interim, and now has a very rich theory and a very rich body of algorithms that basically on the one hand meet this sort of very strong notion of data privacy which has to foresee the possibility of triangulation through the combination of multiple data sets on the one hand but still permits sort of powerful use of data.

And so, you know, one kind of pseudo-theorem that I will state to you is that everything that we pretty much know how to do today in machine learning we also know how to do in a differentially private way. And it's just a matter of companies adopting this technology and choosing to, you know, do their machine learning and data analysis in a differentially private way. And we’re actually starting to see large-scale deployments.

Both Google and Apple use differential privacy in meaningful, large-scale ways in some of their services, and maybe more importantly, the 2020 U.S. Census, every single statistic or report that they release as the result of the 2020 Census they are planning to do so under the constraint of differential privacy. And so this is an example, I think, of a very promising kind of case study, right? Of course, people have thought about different definitions of privacy and data analysis and machine learning for a long time. There was a struggle to kind of come up with the right definition. Many of us believe that sort of definitions based on anonymization are fundamentally flawed.

But then a better definition came along around 15 years ago. There's been a huge amount of research on this particular definition, and, you know, the good news is that in this particular -- for this particular definition of privacy and this particular social norm, it is possible to sort of give these very powerful guarantees at not too great a cost to accuracy or computational efficiency and the like. We can sort of, you know, have the best of both worlds, if you like.

So let me say a few words about research in fairness in machine learning and algorithmic decision-making, which is much more recent. It's a much more nascent field than the study of privacy and machine learning and AI, but we already know a fair amount about it. And one of the things we already know about it is that it's going to be a little bit messier then privacy. So my claim is that if you waded into these literatures and you looked at the work that’s gone on in differential privacy and looked in particular at the definition of differential privacy, you perhaps, like many people, might sort of agree that this is sort of the right definition of privacy.

So we already know that there’s not going to be a right definition of fairness in algorithmic decision-making and machine learning. And what do I mean by we know there’s not going to be a right definition? So there’s already from the last several years several examples, several papers which have results of the following form. They basically say, well, you know, whatever your definition of fairness is, wouldn't we all agree you’d want it to have at least the following three properties. And you kind of look at those three properties and you’d say yes, yes, I would definitely want any definition of fairness to at least meet those three properties and probably other stuff, too.

And, then, of course, the punch line of these papers is, well, guess what, here’s a theorem proving to you that you cannot simultaneously achieve all three of those properties in any definition of fairness. Okay, so those of you that are -- have kind of an economics or social choice background might know about, like, arrows and possibility theorems for sort of voting systems. These have a very similar flavor.

And these -- and so this has very concrete consequences. So in particular, a typical notion of fairness in machine learning would ask for the approximate equality of false positive or false negative rates across different groups. So let me give an example. Suppose you're trying -- you know, suppose you’re a mortgage company and you’re trying to build a statistical model that tries to predict from people’s loan applications and credit history whether they will repay a loan if you give it to them or not, okay? A very natural thing to want to do. And you want this model so that when people apply you can make a prediction about whether they’ll repay or not, and then you want to give loans to people that will repay you and not give loans to people that you predict won't repay you.

But because this is a statistical model, you’re going to make mistakes. You’re going to have both false positives and false negatives, right? And we might think of false negatives as really the case that causes harm to the consumer in question, right? False negative being somebody who's creditworthy would have repaid the loan if you didn't give it to them, but you decided to reject them, okay?

So we might think of false negatives as a harm inflicted on a consumer, and false positives are sort of the people that got lucky. So a typical definition of fairness would basically say that, look, you're going to make false -- you're going to make false rejections. We're not going to try to prevent that, but across different racial groups, it cannot be the case that the false rejections rates differ wildly. It cannot be the case that the rate at which you falsely reject qualified African American applicants is three times the rate at which you falsely reject qualified white applicants, okay? So this is a very natural constraint. And these impossibility theorems basically say if you ask for that and you also ask for a quality of false positives, i.e., the people got lucky, and one other related condition, you cannot simultaneously achieve all of these.

So what this means is that we already know that in fairness we're going to have to simultaneously entertain multiple competing definitions of fairness.

And so not only will there be sort of tradeoffs in competition between fairness and accuracy, there is even going to be competition between different notions of fairness. If you optimize for one notion of fairness or constrain for one notion of fairness, you might be damaging or losing on another notion of fairness, okay?

But nevertheless, you know, we know this and we have to proceed anyway, and so there's been a great deal of research in the last several years on algorithmic fairness and on different notions of fairness and what the tradeoffs between that particular notion of fairness and accuracy is. And this is an area where, you know, to again echo something the previous speaker said, when you sort of talk about the potential interfaces between technical people and policymakers and other stakeholders, I think there are very, very good starting points.

So one thing you can do, for instance, is if you pick a particular definition of fairness like approximate equality of false rejections in a lending application, and you have a data set in front of you, a historical data set of people who did and didn't repay loans, you can actually trace out empirically -- I can give you -- I would have shown this slide if I'd met the deadline -- I can actually show you an empirical tradeoff where on the X axis would be the error -- the predictive error of your model. On the Y axis would be a numerical measure of the extent to which you violated this fairness notion, so 5 percent would mean -- sort of 5 percent unfairness means that let's say between African Americans and other races there's as much as a 5 percent disparity in the false rejection rates. And I can just trace out a curve for you that shows you the menu of choices you have.

So you can get the lowest error, but, you know, if you sort of ignore fairness entirely, that will give you the lowest error but it will give you the worst unfairness. At the other extreme, I can demand that the false rejection rates differ by 0 percent across populations. It's a very strong constraint. And I will get the worst error but the most fairness, and in between you'll get something in between.

And I think this is the type of, you know, sort of quantification of the tensions that we face as a society in making these kinds of decisions that's the right at least initial interface between, you know, people like me and people like you for lack of a better term, right, because it sort of really shows starkly the choices that you have available.

Just to say a little more about fairness, most definitions of fairness, like the ones I’ve been discussing, actually only hold at the group or aggregate level. So you’re only making promises to sort of groups of people in a statistical sense, and you’re not making promises to individuals, so, you know, sort of more prosaically, if you are a -- you know, if you’re a person of a particular race that was falsely rejected for a loan, you would have repaid that loan, the consolation that you have in these types of definitions is, like, well, we’re also rejecting people from other races who would have repaid their loans at the same rate that we’re rejecting people from your race, which is sort of cold comfort if you're somebody who was mistreated in this way.

And so a lot of recent research, including some of my own, is in trying to move towards definitions of fairness and studying their algorithmic implications that try to make finer-grained promises. Maybe not all the way down at the individual level, but to much finer-grained groups than just things like, you know, race -- you know, top-level race or gender or the like.

So these are two social values -- privacy and fairness -- on which in relative terms we know quite a bit already scientifically. And I think we're well on the way to kind of developing both a science and an engineering of designing better algorithms and understanding what the tradeoffs are between accuracy and the various definitions of the social values that we've come up with.

My former grad student and colleague, Jen Wortman Vaughan, is giving the keynote tomorrow. And she's done a lot of recent research on interpretability and transparency of machine learning, which is another, of course, important social norm. I think we know a lot less there so far, partly because we just haven't had as much time, but one of the problems with sort of coming up with satisfying definitions of things like transparency and interpretability is that there's fundamentally an observer in kind of the middle of such a phenomenon, right? So when you talk about interpretability, for instance, of a statistical model, you have to talk about interpretability to whom and for what reason and in particular the sort of numeracy of the audience you have in mind will matter greatly, right, if we're talking about interpretability to people with like a statistics training, that means one thing. If we're talking interpretability to the average American citizen, well, you know, the average American citizen has not been exposed to linear regression and may find it a little bit bewildering to even talk at all about an abstract mathematical mapping from loan applications to lending decisions.

And so I think much of the research that needs to happen on that topic will have to have like a cognitive and behavioral element to it. You'll need to do human subject studies with the type of audience that you're interested in and ask them what they think is interpretable to them or whether you can explain models to them and the like.

So I'm almost out of time, but just to sort of quickly touch on a couple of other things that I saw on the agenda, I saw that there was one discussion -- there was one panel title that had a very intriguing name, which was Algorithmic Collusion. And I'm not sure exactly what the context that's meant there is. But, you know, if your concern is that, you know, we might be entering an era where algorithmic decision-making causes in some implicit or explicit kind of large-scale way collusion between different entities, whether it's on things like pricing or decision-making and the like, I definitely think this is already happening.

One area in which I'm very familiar with this already is on Wall Street where quantitative trading teams tried to build statistical models to predict the directional movement of stocks and, so to speak, beat the market. And my basic belief there is that there’s a huge amount of implicit sort of collusion going on there, and it's really because, you know, when we all use the same or similar data sets, and when we all use the same or similar algorithms to train our models, then even if we think we're clever and independent and creative, we are going to be strongly correlated just through the data, right?

If we’re trying to predict the same thing and we’re using similar data sets and similar methods, then no matter what else we do -- everything else we’ll do is a second-order effect from the fact that the data itself will correlate us. And so I think that this is an interesting topic on which there is probably interesting scientific things to say but I haven't thought about it yet, and I don't know of a large body of research on it.

But I'm out of time, so let me stop there and let the agenda move on.

(Applause.)

DR. GOLDMAN: Well, thank you so much, Professor Kearns, for that great overview and introduction to all the issues that will be covered in this hearing.

DR. KEARNS: Okay, thank you.

DR. GOLDMAN: So now it is 10:15, and we're going to be taking a little break until 10:30, at which time we will be back for the first panel.

(End of Presentation.)

UNDERSTANDING ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS THROUGH REAL WORLD APPLICATIONS

DR. GOLDMAN: So we're now going to begin the first panel on Understanding Algorithms, Artificial Intelligence, and Predictive Analytics through Real World Applications. As I mentioned, in case someone had just come in, I'm Karen Goldman. I'm an attorney advisor in the Office of Policy Planning at the Federal Trade Commission. And this is my comoderator, Dr. Harry Keeling, who is an Associate Professor in the Department of Computer Science here at Howard University.

So we hope that this panel, which will cover applications that are currently in use and on the horizon, will provide a sense of the variety of uses to which these digital tools can be put and highlight that no single application is truly representative of their use.

I just want to mention that anyone in the audience who would like to ask questions of panelists should write their questions on the notecards that are being passed out and will be collected later on.

With that, I'd like to introduce the distinguished members of this panel. So we have Dana

Rao, who is an Executive Vice President and General Counsel of Adobe. Next, we will have Henry Kautz, who is the Division Director for the Division of Information and Intelligent Systems in the Directorate for Computer and Information Science and Engineering at the National Science Foundation.

Then we will have Angela Granger, who is Vice President of Analytics at Experian. And then Melissa McSherry, who is Senior Vice President, Global Head of Data Products at Visa. We have Michael Abramoff, who is the Founder and CEO of IDx, and Professor of Ophthalmology and Visual Sciences at the University of Iowa and also Professor of Engineering and Computer -- of Electrical and Computer Engineering and Biomedical Engineering.

And then we will have Teresa Zayas Caban, who is the Chief Scientist at the Office of the National Coordinator for Health Information Technology.

So with that, Dana, would you like to begin your presentation?

MR. RAO: Thank you.

Thanks for being here. So the first thing I wanted to just sort of get out there, I'm a lawyer, and people are like, why are you talking about AI, and

I thought I’d put it out there because there are some very distinguished computer scientists on this panel. So I was actually an electrical engineer undergrad and going to a university, and so when I was at law school, I was going to write a paper, a note for the journal, and this book was on my dad's desk, *Understanding Neural Networks*. This was back in 1996. So I thought, oh, that would be fun to read, and I read it, and I wrote my paper, which got published, called "Neural Networks -- Here, There and Everywhere,” which was a wildly inaccurate characterization of where neural networks were in 1996. So don’t come to me for your stock advice, but it was -- it’s been a fascinating topic for me, and at Adobe, we're really interested in this topic.

And for us, AI is special because we have this entire business that’s focused on helping people be creative. And creativity is a part of the brain that doesn't follow rules. It's unstructured, and traditional software programming is a very structured form of algorithms. It’s predictive. You understand the rules, and you understand how to characterize it, and that's actually not a great fit for creatives who tend to break rules.

And so our products have always struggled to bridge that gap between innovation and creativity and the structured form of traditional computer programming. And AI bridges that gap, and it really allows us to create tools that are better for our creative customers.

So when we think about how we look at AI and digital creativity, we're really focused on minimizing the mundane, eliminating those repetitive tasks that everybody has in their day. And so for creative professionals, there’s a lot of complexity in the tools and in setting up the camera shots or the video shots that are not actually the highest value added that they have, where they’re really trying to get their artistic sense across or fulfill the goal of a marketing campaign as they create content for it, the complexity of adjusting each pixel’s luminance or the color or the background or the lighting gets in the way of them actually doing the part of the work that they're getting paid to do. So that's where we're really interested in using AI, so it would eliminate those mundane tasks.

And we also at Adobe, we’ve noticed there’s a huge demand for content now, and that's either because there’s social media channels and people are posting content all the time on Instagram and Snapchat and Facebook, or on ad campaigns -- digital media advertisement campaigns where -- digital marketing campaigns where you are personalizing content for each consumer. So there’s a huge demand for content, more than ever before, and our creative professionals need to be able to create content at a higher velocity, and that's what AI is helping us do.

So when we think about AI, we think about it in the creativity space in two buckets -- content understanding, computational creativity. And Adobe also has an experience intelligence business. I’m not going to talk about that much today, but just for transparency, we also have this other business that also uses AI in a different way.

Content understanding is really trying to get behind what’s in an image, for example, or a video. So it's easy to look at an image of a cat and say there’s a cat, or there’s a house and just do sort of basic object recognition. What AI allows you to do is provide that insight into the image and add an abstract layer, a conceptual layer above what you typically can do pre-AI so we can understand things like actions and concepts and styles and sentiments, so just abstract concepts that are in your image that the AI can infer from looking at it.

So we have a couple of demos that we’re going to show. We're hopeful they’re all going to work correctly. I think this is going -- yeah, it’s going. And these -- this deck will be published in the Adobe public policy blog, so anyone who wants to see the full deck and watch the videos through can do that. But we’re just going to talk through a few -- a couple seconds of these.

So this is a person in the, let me just go back here. Set this up. So this is a person using our stock photography service. And so they wanted to start a creation. And so they wanted to be able to say, I have an ad campaign for Nike, how should I start. And they go to our stock photography site and they just search for things to sort of -- as inspiration for the ad campaign.

And so for example, in this example, this person’s going to say, you know, I see this image of this woman with a ribbon jumping. That sort of captures the aesthetic of what I want. And here we go. And so she -- say they choose this picture, and then what Adobe Stock does, it recommends other pictures that are very similar to this picture. So in this case, she says, okay, I like this, this is a good start for me, and then Adobe Stock at the top does sort of a normal picture recommendation. Here are other pictures of people with ribbons, and that may be what you're looking for.

But in this case, that's not what we want. Like, Nike actually wants this sort of freedom. And so we select the woman jumping, and our AI understands that what we want is actually the action of jump. Like that’s what we want out of this picture, not the color, not the ribbon, not the blue sky. We want the action of jump. And so now we actually recommend pictures that are about jumping.

So we can take the concept of that picture and using AI understand, okay, they actually wanted jumping, and so now we can just show them these other pictures.

Now, the next level is we say, okay, well, Nike didn't really want a picture of random people jumping. It was actually supposed to be a family picture. So we take family and we use the jump concept from the first image, so you see how they’re stacked on the right, and now you have families jumping. And now the creative professional could say, that’s where I want to start, I want to choose one of those pictures and start my campaign from there.

So how do we do it? So what we do is our AI will analyze these -- in this case, an image -- and look for the concepts behind it. So you can see in the middle, there’s concepts, and on the right, there are percentages. The percentages are the confidence that our AI is actually accurately predicting what is going on in there. But what you can see is we’ve analyzed those faces and we’ve analyzed the context of the picture, and you can see that where you said, oh, there’s happiness there, there’s love there, there’s joy there, we’ve understood the abstract concept of those pictures. And so you can go, if you're a creative professional, and say I need pictures, my theme is love, you can type in love as a search term, and you’re going to get a wide variety of images, but they’re going to have this concept in them.

You can also look for families, right? And it will understand that the connection of these three people plus the expression on their face means that they're a family. And you can understand -- and you can search for concepts like family as part of this. And so you can see all the different kinds of categories that you are able to search on using our Adobe AI technology to analyze what is actually going on inside the picture.

We also have a PDF and Acrobat service, you know, and that has lots of text, and we’ve actually run our AI on the text to understand the intelligence behind the words. And we have married that up to images to allow you to do automatic phrasing. And, again, we can do very basic captioning. So you put your photo there, and we can say couple on a bike and that’s object recognition. But then we use AI, then there’s a little slider you can see that’s moving. And you can say I want to see what the AI thinks this is. And it says young couple on a bike, or in this case, it said beautiful peacock, right? So it understands not just the image but also the concepts behind the image. So if you wanted to search for “beautiful,” you’d get that peacock, for example.

So these are the techniques that are being used when we talk about content understanding, the first part of how we looked at AI and creativity. You know, it’s traditional machine learning, it’s traditional deep learning, and we look at all these things like aesthetics and style and color, we train our AI to understand these concepts, and then we are able to provide these services to our creative professional.

The second piece of what we do is try to make the creative professional day's faster. And that’s what we call computational creativity. And that is trying to help their work flow. How do we help them do those tasks even faster than they used to have to do under traditional software? So here’s an example. Let's say somebody wants -- Macy's wants an ad campaign and they told you to go out and shoot a cityscape at night, and you go out and you spend six months getting this shot. It had the right lighting, the right building, the right angle, and you’re like, all right, I'm great, I’m happy.

And then you turned it in, and Macy's was like, you know what, we’ve changed our mind, we want a different setting. We want it to be the sunset. And so then traditionally, you’d have to go spend another six months reshooting this picture trying to get the lighting correct.

So with our AI, we can automatically segment out the part of the picture that’s of interest to you, and then that's the cityscape. And then we let you import another picture that is of the desired lighting and sky that you want. And with one click, you can now take that lighting and put it in your picture.

So that's probably not 100 percent of what the creative professional wants for their Macy’s campaign, but it's probably 80 percent or 90 percent of what they want, and now they can take this picture and make it into exactly what they want with very little extra effort. So you've just taken six months of extra work, of not exciting work, that was not the fun part of their day. The fun part of their day was setting up that shot to get that image in the first place. And now they can take this and they can go back to Macy's, and if they come back and Macy's says, you know what, we've changed our mind, snowy, blue-sky day, five minutes later, you can just change. And so the AI really helps drive that routine out of your day.

Another example is what we call neural stylization. And so, again, this is the idea that we've been able to understand the style of an image. And so we've trained our AI demonstration on the style of different famous paintings. And so if you have your photograph on the left and you said, I want it to look something like the interpretation of these two different paintings, you can do it. All it does is understand the style of whatever painting you put in, and it's just the style of it. So it's not just copying the colors broadly like you might have expected pre-AI. It understands what the style of the image was and applies it to the image.

So just understanding that concept of -- I think this is going to play. And so this is not just for creative professionals. This is for hobbyists. You can take your own pictures and you can upload whatever artist you want and it’s going to take the style of the artist and apply it to your picture. And it understands that concept.

We can also use AI and we do use AI for our video editing products. So this is a project called Project Cloak, and this is a normal example where you have -- you shot a scene and then in post-production, you want to get rid of something you don't like. In this case you don't want that couple there. So using AI, we segment the image and understand who’s in the image and who they are, and we can also fill in the background with copied pixels to make the background look perfect.

So on the left is the original footage, and on the right is post-AI, and it looks like they’ve just vanished, right? And then that used to take months of work to do to edit two people walking out of the video, and now you can do it in minutes.

So as I mentioned, we also have an experience intelligence business. This is the other side of our business. This is a digital marketing business that allows you to target advertisements and allows chief marketing officers to understand what the content in the campaign is doing. So we provide that service and use AI there as well. We use it to help you predict the results of a campaign before you even launch it. We may say this is going to be successful in the northeast, or this is going to be successful in California based on our analysis of customer data from interacting with their website. That's another way we’re using AI at Adobe.

So I think the question is how we get there. How do you actually produce the AI, and I know there’s going to be a lot of people talking about the nuts and bolts of the computer science so I’m not going to spend too much time on this, but this is how Adobe does it. Our AI product is called Sensei. And this is the architecture.

And so what we do is -- what we do typical of any neural network, we have the neural network and then we train it with data, and we train it for an outcome. And using this architecture, we're able to create the neural network; we freeze it in place; and we ship it PhotoShop; we ship in Premiere, and that's the result you see as a consumer.

So the principles -- this is my second-to-last slide -- the key principles for training AI that is important to Adobe and just a takeaway for everyone is how do we make this product work well is we need millions of pieces of data to train it. You need lots of examples of artists; you need lots of examples of images in order to train a neural network to understand the insights that we’re able to show you.

So when you think about how do we make AI beneficial, how do we get the rewards of AI, you need access to data, you need access to a lot of data and you need access to a variety of data, and that variety of data will make your neural network accurate. And a variety of data will also eliminate bias.

You can imagine bias when you’re looking for images, that is inherent because you may have trained your AI on a particular kind of a person, and if you go searching for a job or an occupation, you’re always going to get that person because that's what you trained it with. So the wider variety of data you put into the AI, the more likely it is your results are going to be unbiased.

So thank you for your time. This is our presentation. Creativity in AI, with AI is what Adobe is focused on. It's how we believe AI will help transform the creative professionals for today and tomorrow. Thank you.

DR. GOLDMAN: Thank you so much for that colorful and creative presentation.

So, next, Henry Kautz will begin his presentation.

DR. KAUTZ: Thank you. So I’m going to focus my talk on the work we’ve been doing at NSF to support AI applications for social good. So when we look at a proposal, we have two major criteria. First, we want to advance science or engineering, looking at fundamental advances, but we also consider potential for broader positive impacts on society.

Now, the traditional broader impacts that were frequently mentioned in proposals, we’re training graduate students and potential future applications of the result. So someone, say, I'm doing this fundamental research, and maybe someone in the future will come along and do something to benefit society with us.

But, increasingly, we see that the fundamental science and these broader impacts are entwined, that as you work on an application for social good you discover new questions that require fundamental scientific advances. And from those advances, you discover that there are new opportunities.

So AI and broader impacts. So AI methods, taken broadly, that includes machine learning, knowledge representation and reasoning, and what we might call deliberative intelligence, making optimal decisions, are being used by researchers in every discipline that’s funded by NSF. I'm from the computer science, and my particular division funds a lot of the fundamental work in AI, but there’s really no area of NSF now, including the social sciences, where you don't see people talking about AI. And, increasingly, we're partnering with other agencies, that are funding or taking advantage of work in these fields.

So we’ve seen -- over the last decade, we’ve grown up quite a rich portfolio of what we call cross-cutting programs. So these are interdisciplinary funding opportunities that involve multiple directorates within NSF and sometimes with other agencies. Some of the most important are the Smart and Connected Health Program that we run with NIH. And so there, we are looking at AI research that is a bit more applied than traditional work funded by NSF but is not yet ready for the kinds of actual clinical uses that NIH would fund. So we both put money in there, and then we help bridge the gap between those agencies.

Smart and Connected Communities looks at applications of AI to all kinds of problems facing urban life from pollution, policing, and violence, transportation, other issues. We’ve had a program for several years now called Big Data in science and engineering, and that is to support broad collaborations -- collaborations that can cover a number of fields. So you might have material scientists together with a computer scientist or, you know, electrical engineer together with the computer scientist or even medical people.

And through that Big Data program, we’ve also funded what are called big data hubs, so the idea that these are a set of universities that act as resources to all of the universities in that region for activities such as helping making connections to government agencies. And through that, we’ve had programs like the Civic Innovation Challenge.

One of our most recent programs that is particularly relevant for broader impacts is one called the Future of Work at the Human Technology Frontier. And it's a very interesting combination of directorates -- computer science, engineering, education, and then our social, behavioral and economic sciences.

So we’re now looking at the future of the workplace and in particular how AI will be impacting that future. So we want to fund both the kind of technology we might see in the future. So, for example, in a recent -- we just completed the first year of the program, and one of the awards was on smart classrooms, so how we might integrate AI as a teacher’s assistant, and not replacing a teacher but assisting a teacher.

But we also will be looking for work where technologists work with social scientists to look at both the positive and the negative consequences. Will AI throw millions of people out of work? That's absolutely an open question. If you look back at the history of science and technology, you can make quite good arguments either way that AI will lead to permanent unemployment or that AI will lead to new opportunities for employment.

This is another example of the work from this most recent program solicitation. So Whole-body Exoskeletons for Advanced Vocational Enhancement. So, here, we’re looking, you know, at something a little bit different than your traditional robotics for manufacturing but augmenting the human worker to give the human worker superhuman strength and endurance, or as I mentioned in classroom teaching, where a system that is monitoring a classroom and noticing when students -- those students who have become apparently disengaged are not working or not attending and can perform such tasks as simply alerting the teacher or engaging in a personalized activity with the student.

So one of our very largest grant programs is called Expeditions in Computing. These are typically $10 million over four to five years. So, here, we’re really looking for research of the highest intellectual merit. All of our reviewing is a system called peer reviewing, where we get unbiased scientific experts from the community to review. And in Expeditions, we have multiple layers of reviewing because we really want to get the best of the best.

And in addition, these -- the work we fund should address the nation's greatest needs. So to give just a case study of the synergy between positive broader impacts and scientific merit, I’d like to just mention some of the work going on at the Institute for Computational Sustainability, which is a -- the result of actually two successful Expeditions in Computing that went to a consortium of Cornell, Stanford, and

University of Southern California.

So the problem here is looking at sustainability problems, and by sustainability, we're looking at environmental sustainability, economic sustainability, resources, social sustainability, very broadly, as complex problems that are really too difficult to solve with human intelligence alone. So we want to employ AI techniques and large amounts of data to solve optimization -- essentially resource optimization problems that are far beyond the kinds of linear optimization that most of the people in this audience would be familiar with.

These are highly nonlinear problems where we must model uncertainty. So we can't -- we just can't ignore the fact that many -- there are many variables that are unobserved. Okay.

Now, you might think that, well, these are all different problems, but what has been so fascinating by this Expeditions is that problems that seem to be quite different often have very -- have shared technical solutions, okay? So this is a subway map that the research group created. And as we see, each of the tracks of the subway, the six tracks -- the six tracks are scientific themes. So pattern decomposition, crowdsourcing, mechanism design, so social choice theory, and economics, spacio-temporal modeling, probabilistic inference, and sequential decision-making. And then each of those tracks is going through the stops, where the stops are the particular application.

So in each application you had domain experts. So let's say there's one there on landscape scale conversation and rural communities. That included, you know, people who knew a lot about that topic and had been studying and working with communities in Ecuador, but it made use there of temporal modeling, probabilistic inference, and sequential decision-making. So you see it's quite a variety here -- flight call detection, and I'll mention that again, wind and solar forecasting, all the way over to microbial fuel cells.

Now, but one thing I should point out is AI covers many things. There's sometimes a tendency because of the great success of what are called artificial neural networks to say that that is AI. And as we just saw from the previous speaker, artificial neural networks are wonderful when you're dealing with patterns, doing pattern recognition, and essentially trying to emulate those parts of intelligence that don't involve essentially logical thinking but are more based on pattern recognition and intuition, the kinds of problems we don't think about when we solve them -- recognizing your friend’s face, right? We don't think consciously about it.

By and large, the work in this particular set of projects, though, involves what we might call your Type 2 intelligence, your deliberative rational intelligence where you have many alternatives to consider. In fact, there is such a large number of alternatives, you can't simply enumerate them all one after the other. You have to have very clever ways of essentially searching through an enormous, sometimes infinite space of possibilities and narrowing in on those points that are near optimal.

So just going down a little bit deeper, the problem of data -- of decomposition in big data. So this is -- so a core technical problem. You have some kind of very complex signal, and you want to reduce it to something simpler, right, to a small -- the one measurement or a small number of measurements. So this is also called dimensionality reduction, source separation, and segmentation with complex constraints. But it makes use of a body of algorithms that have come up in computer science, electrical engineering, and particularly more and more in work in AI.

So we had a -- there were a series of projects, one on detecting gunshots. And you can imagine security applications in a city. Another on detecting elephant calls. So you can put out audio monitors in the jungle and use that to conduct a census of elephants, right, based on their calls. That same work was then used to detect birdcalls of actually birds in flight for a project with the School of Ornithology at Cornell. And perhaps, surprisingly, is with very few changes, that same algorithm was used in a project on crystal phase mapping, which is in material discovery, so a problem where you're coming up with a mix of new materials, you hope they have some property, and you're analyzing the results of shooting x-rays at those new materials.

Another example -- my last example here -- is dealing with hydropower in the Amazon Basin. So there are a great potential for getting more hydropower from the Amazon Basin. And, in fact 170 dams have already been built or under construction, and about 300 dams are planned or proposed.

Now, there's obviously a big problem here. If all of these dams are built, not only will there be quite a lot of devastation to wildlife, but they will become less effective because one dam is going to affect the water flow to another dam. So you have to look at this as a multi-objective optimization problem to balance off energy, fisheries, transportation, and navigation. Obviously, as you put in more dams, you’re going to make river transportation more expensive, and finally looking at the long-term effects, how will all these dams affect the natural flow of sediment and nutrients and how that affect farming. So this becomes a multi-objective optimization problem.

And then the goal is to look at the tradeoffs between these different factors and have a new algorithm that can present, well, here is the possible best tradeoff. There’s no single best tradeoff, but you can look at that any solutions that don't fall along this line are provably worse, so they're worse in some respect and no better in any other respect. So this tremendously reduces this sort of infinite space of the number of dams and the placement of dams to one that now can be decided by humans. Yeah, that’s showing where they’re the dams.

And interesting that this same effort has led to startups. For example, ATLAS AI, that is basically a for-profit AI for social good company. This also received funding from the Rockefeller

Foundation, looking at providing -- helping developing nations be more sustainable in their agricultural practices. Networks of CompSustNet, a larger network that includes this group of these three universities with others to address these kinds of problems.

And with that, I’ll conclude. Thank you.

DR. GOLDMAN: And thank you so much for showing us the diverse portfolio that NSF is supporting.

And, now, Angela Granger will begin her presentation.

MS. GRANGER: Thanks. Sorry, it’s a little tight up here, so we thought that would be the better route to get around.

I lead analytics for Experian, and one of those areas that I'm responsible for is credit scoring product development, and for those of you that don't know, Experian is a global leader in consumer and business credit reporting and marketing services. We support clients in over 80 countries, and we have approximately 17,000 people in 37 different countries.

We believe it's our responsibility to assist lenders in managing consumer credit risk and empowering consumers to understand and responsibly use credit in their financial lives. We’re committed to being the consumers’ credit bureau, and I thank you guys for having me here today.

To set the context for today, there’s a lot of different areas of application for credit scoring, so we’re going to -- I’m going to specifically talk to scores used to assess eligibility for credit where adverse action may be taken. The example was used a couple of times earlier today specifically of application of credit for an example where you could be approved or declined, your application for credit. That would be the credit scoring context we're talking about today.

Benefits of AI or machine learning, for both lenders and consumers in our industry, are ultimately better lending decisions. If you have greater insights into the data that you're using, better accuracy in the scores, you’re going to have better decisions being made.

And, secondarily, financial inclusion. Where we’re really finding the power of AI and machine learning techniques is our ability to evaluate new data sources more quickly and incorporate that new data into credit scores, thus broadening the access for credit for people who maybe have thin credit or are new to credit and don’t have a credit file with us today.

Where we like to start is with the data. If you think about predictive modeling, and any kind of modeling for that matter, it's important to understand the data that’s feeding into the model. For us, we talk about traditional credit data. And when you think about traditional credit data, what we refer to is what you typically find on the core credit databases at the major credit reporting agencies. And this includes information around what we call trade lines or account-level information where you get access to a consumer’s payment history on a certain type of account, their outstanding balances, that sort of thing.

We also have information on inquiries that are made into the credit bureau for applications for credit as an example. And we have public record information, particularly on bankruptcies. We also maintain some additional information that you might think of as being part of a credit application, such as income and employment.

We also like to talk about alternative credit data. So this goes by many terms. In our industry, when we say “alternative credit data,” we really mean data that is not on that core credit database that I talked about a minute ago. So types of alternative credit data that aren't reported to the core credit database today can include rental payments, asset ownership, alternative financing such as payday loans, short-term loans, rent-to-own-type loans.

There's additional public record information out there that's not on the core credit database. And, most recently, we've incorporated consumer permissioned data.

Both alternative data and traditional credit data have been found to be very predictive of a consumer's creditworthiness. And, particularly, the alternative data comes into play in those cases of thin file and no-hit-type consumers that I mentioned a minute ago.

The Fair Credit Reporting Act regulates the collection, dissemination, and use of consumer credit information, and so all data used in credit scores are what we would call FCRA-complaint. What does that mean? That means the data needs to be accurate, so the credit reporting agencies must do their best to ensure their data is accurate. The data is disclosable, so consumers can see that information. Consumers can get one free credit report every 12 months, and they can see their credit information if they're denied credit as an example.

The data furnishers also play a role in the process. They have to confirm information when disputes happen, and they're held to certain turnaround times as well as part of the dispute process. And, lastly, we were set up pretty nice earlier around fairness. Fairness is another part of the FCRA. So scores are -- they cannot discriminate based on these different ECOA factors such as gender, marital status, race, and religion.

So for about 30 years, we've been using scores kind of in their current form, which means they're using this core credit information that I talked about earlier. And so between that and our experience over time, we've come up with things that are generally acceptable in our space, data that complies with those FCRA rules that I mentioned earlier, proven payment information, rental data, account transactions from your demand deposit accounts are generally deemed acceptable. Generally not acceptable are things like social media data, you know, who your Facebook friends are sort of thing, and any data that could discriminate in decisions or that could be discriminatory, I should say.

Under consideration right now, we're looking at education level, again, something to help us in that new-to-credit space. Think of students graduating from universities and having that information available so that they can more easily get credit and join the credit ecosystem.

So one of the things about our industry is not only is the data itself, which we just went through, regulated but the scores or the models are regulated as well. There's regulatory guidelines around accuracy and fairness that have been put out by the OCC. Those documents or those guidelines, I should say, are pretty extensive. They cover the model development process, they cover model use, they cover model monitoring, when to redevelop. And they create quite an extensive amount of documentation.

And in order to meet these model governance guidelines, many of our clients -- so think of, you know, big banks, big lenders -- have had to create entire staffs just to take on this model governance requirement.

We talked about the controls around discrimination which lead to the need for transparency. And then in the FCRA, we are also required to provide your top four reasons for your score being what it is as well. And so the need for transparency, or what we call explainability in scores, is very high.

Some key considerations when developing credit scores to meet all these needs, I won't go through all of these in particular, but they really cover the full life cycle. We talked about, at one of the earlier sessions, generalization. So our models need to essentially replicate. They can't just work really well on the training sample. They have to work well in production. If you think about credit scores in use today -- think about mortgages in particular -- the credit scores being used there are about 20 years old, right? So these models need to continue to replicate and still rank-order consumers in terms of their creditworthiness.

Today, models have an average shelf life of about three years, so we're looking at AI to help us get models to market faster. Some research that we did, we tested several different techniques around machine learning. I won't go into each of them. You can see that here. But suffice it to say the gradient boosting models are the ones for credit scoring and credit risk in particular that seem to be rising to the top.

When we let the machine run by itself, these are the type of results we get. We see anywhere between a 5 percent to 10 percent lift depending on the situation. This is a more generic sample for auto and bank card, so we see about a 5 percent lift if you were to do the math here. But our clients report anywhere up to a 15 percent lift as they start to really look at specific portfolios or specific lenders.

This, however, is when you just let the machine run itself and you don't take into consideration some of those things we talked about earlier.

We do something that we call model refinement, and this is where you have to go in and ensure your model is working as expected, that you can explain what’s happening. You want to make sure that a credit score doesn't go down if a consumer has made some impact to their credit such as paying off some of their debt or lowering their utilization. And if you don't do this refinement and you don’t understand what’s happening under the covers, that can happen.

So when you go in and you refine the model through the requirements that we talked about before, you’ll see that the lift in performance from the -- in this case, extreme gradient boosting methodology, is lessened. So in our particular example, the lift went from 5 percent to 2 percent. In other examples, we've see that 15 percent or 10 percent lift come down to 5 to 8 percent, right? So on average, we're seeing about a 5 percent lift in accuracy from applying some of these techniques outside of our traditional regression methods.

This is just another example of addressing overfitting, which tends to be a problem with some of these new methodologies that aren't -- haven't been used in practice as long. What you tend to do if you throw all of the data into the machine and let it do its work, we have over 2,000 attributes, variables, characteristics that we will throw into a model, and it will use almost all of them if it can, right.

And that tends to overfit and the model doesn't generalize. And so you do have to go in and manually intervene and not let the machine do all the work.

Some of the advantages for AI in credit scoring go beyond just the modeling. You know, I mentioned a 5 percent improvement, and I'm sure you guys are all sitting there, going, whoo, 5 percent, 5 percent, right? But in the credit risk world and creditworthiness world, we have very predictive models today. And so a 5 percent improvement is actually a big improvement. The data that we use in the models is very accurate, and so we get very good models. So 5 percent improvement is significant, but we're looking to use machine learning and AI methodologies across the model development life cycle and not just in the model development itself.

Probably most importantly to take away from today is in credit scoring. Credit scores are static models. So most of us when we think of AI think of realtime updating, self-learning type models. Those are not in use in our industry today. These are static models. So while we're looking at these additional techniques outside of regression, we're still talking about static models. I mentioned the turnaround time or the shelf life of a model is about three years right now. With these new techniques, that's going to come down, but we have to have the ability to go back in time and replicate our models.

So, lastly, there's some future policy regarding credit scoring that we wanted to make sure you were aware of. Today, unlike what people think, your telephone bill, utility payments are not reported to the credit bureau. Those are very powerful predictors just like other payment methods of future payment behavior and so of creditworthiness. And there's been several studies that show that today.

And so with that, I would like to thank you for giving me this opportunity and hopefully this gave you a quick glimpse into the status of AI and how it's being applied in credit scoring. Thank you.

(Applause.)

DR. GOLDMAN: And thank you, Angela, for that very interesting presentation on credit scoring and bringing in the related legal and policy issues.

So, next, Melissa McSherry will begin her presentation.

MS. MCSHERRY: Thank you very much, and thank you so much for having me today. I work with Visa. Visa is the world's largest payment network, and what that means is basically when you use a Visa card your -- the merchant where you use the Visa card basically calls their bank and says can I authorize this transaction. And then Visa connects the merchant's bank with your bank, who says yes or no, that's a good transaction to authorize. And then that signal goes back to the merchant, and all of that happens if everything goes according to plan. All of that happens almost instantaneously.

In that -- in that context, Visa is very -- we work very, very hard to make sure that the transactions that are going through are legitimate transactions or not fraudulent transactions. I think fraud worldwide today is something like $600 billion, so it's a lot of money, and we want to make sure that we do as much as we can to help banks prevent any of those fraudulent transactions from going through while still making sure that all of the good transactions go through. Basically, when you are actually the one using your card, if you try to use it, that it actually works.

So what I’m going to talk about today is one way in which Visa is using AI, specifically computer vision, to help us do that work of looking after and preventing fraud on the Visa system.

So you might be asking what do puppy dogs and blueberry muffins have to do with preventing fraud. And I put this up just to sort of illustrate both the challenges and the opportunity in computer vision. So all of you could look at these pictures and very easily discern what’s a blueberry muffin and what’s a puppy dog. But using the techniques that were available up until, you know, call it 2012, 2013, this was actually a pretty hard problem for most computers to solve. They would get it right about 75 percent of the time.

And in I think it was 2013 -- there’s a competition that is run every year. And new techniques, specifically things called convolutional neural networks, started coming into play and started dramatically improving the ability of computers to correctly differentiate the muffin from the dog. And so we're now at the point where these techniques can generally differentiate not just muffins and dogs but can differentiate different images about 97 percent of time as opposed to 75 percent of the time, which is really quite good.

If you think about human beings -- although if you’re sitting there concentrating, you know, you would always be accurate since most people don't concentrate all the time and they do sometimes make careless errors, human beings run at about 95 percent of the time, right, when you give them a lot of images. So this ability to look at a picture and to say this picture looks like this one, and this other picture looks like this other one, this is one of the applications of AI that has dramatically improved. And so now I’m going to talk a little bit about how we use that application of that computer vision application of AI in the context of fraud.

So just a couple of terms before we get started with this particular example. First of all, what is a fraud score? Like I said, whenever you use a card, Visa basically attaches a score to the transaction that goes to your bank that says how likely is it that we think that this is actually you using your card versus someone who's trying to commit fraud using your card. We provide that information to the bank so the bank can make a decision about whether or not they want to authorize the transaction.

And as you can imagine, we process a lot of transactions, right? So that first thing we do in every transaction is we attach a score from zero to 99. But then if we look across all of the transactions, we can actually say, for instance, all of the transactions in an hour, how many of them were at, like, the highest score, got a score of 99? How many of them were at the lowest score, got a score of zero. And it's helpful to us to look at the percentage of scores that are in each of those bands.

And the reason why is if you -- if we're running along and 1 percent of the population is getting the highest score, that 99, and it's nice and steady and then all of a sudden like 10 percent of the population is getting a 99, that means that probably one of two things is happening. Either there’s a giant fraud attack, and there are fraudsters that are trying to, in a very coordinated way, steal a lot money, and this does happen sometimes, right, in which case we need to intervene. And we typically intervene by calling the banks that this is happening to.

Or there is something wrong with our models or system or how we’re processing things. And, again, that's a situation in which we need to intervene and we need to make sure that everything is actually working as we expect. So not only do we look at the fraud scores, we also look at the distribution of those scores.

And so in the next page, this is just -- this is a made-up example, but I think it sort of illustrates what’s going on. So you can imagine that this is a graph looking at the percentage of transactions in a particular score band. And in this particular case, I just did it over days, and it goes up and down, and it goes up and down because, for instance, the kinds of -- the mix of transactions that you see on like a Friday night can be pretty different than the mix of transactions you see on a Sunday morning. And so the mix of transactions in a particular score band can go up and down.

Now, if you look at this, it's pretty easy, again like the puppy dogs and the muffins. It's pretty easy to see that at the end there’s something that looks a little bit different, right, that doesn't -- that pattern doesn't look like all of the other patterns that came before it.

And this is, again, pretty easy for everyone in the audience to see that that’s different, but it's actually kind of hard for the tools that we had prior to those computer vision tools to pick this up, like you can’t -- like a traditional control chart, it’s hard to write a rule that will get this because the actual numbers are sort of -- they’re inside the range of the historical range, they’re going up, they’re going down. They’re not -- it's just -- it’s hard to write the rules. But, again, it's easy to see it using computer vision tools.

And so what the computer vision tools let us do is basically do what a person would do in terms of looking at this and seeing a pattern that's different. But the computer vision tools let us do that every hour of every day. I mean, the computer doesn't get tired and people do, like, they need to go do something else other than look at charts all day.

It lets us look at hundreds of metrics, not just one, right? And if you think about this, this is a pretty simple chart that I put up here, right? This is basically one-dimensional, right? We sort of look at the scores, versus one-dimension. And so it's easy to see the variation. If I had put a chart up here that had multiple dimensions, like we were varying a couple things at the same time, that very quickly gets really hard, even for people, to see the differences. But, again, the computer vision techniques that we've been talking about can pick those variations up pretty quickly and can identify those. So we can not only monitor what's going on versus one dimension, we can monitor what's going on versus multiple dimensions, and it makes our monitoring that much better and that much faster.

So just a quick explanation of how we've applied this in our particular situation. Basically, we built a model that looks at the distribution of each of those score bands that we just talked about, so, you know, for instance, scores of 10 to 19, right, so it does this for each score band. And it looks at those distributions for a five-hour period over each of the last 120 days. Right, so this is lots of data that's coming in. Think of the computer as looking at a chart, an hourly chart over the last 120 days.

And from that, it forms an expectation of what the current five-hour period is going to be, right? Is the score -- is the distribution going to be going up and then down? Is it going to be going down -- you know, down and then up? Is it going to be going, you know, one direction -- it forms an expectation. And, then, and this is the part that relates back to the puppy dogs and the muffins, it looks at the actual picture and it compares it to its expectation that it created based on the last 120 days, right?

And so on the top row, we see on the right is sort of what we would expect, right, for this time period from the data that’s come in over the last 120 days. And what we see on the left is what actually came in. In those two pictures, the computer would say, yep, those two things -- they look similar, they're both blueberry muffins or they’re both puppy dogs, right?

But on the lower band, what we see is the expectation for the particular time period that we're looking at is just that the scores will be going up during the time period. But what we actually see is that they're going up and then coming back down. And the computer at that point says, no, no, no, these do not look like they're the same. This is not -- something is not matching here.

And that, then, causes the system to generate an alert and say, hey, a person, a human being, needs to go look at this, right? It might be that it's fine. It might be that it's just, I don't know, Black Friday, right, and so all kinds of things are different. Or it might be that there is an actual problem and we need to get engaged and figure out what the problem is, and we need to figure that out promptly.

So in this particular case, what's going on is the computer is basically taking a lot of work that might have been kind of boring and tedious for the people and doing the boring and tedious part and then just pulling out the things that are interesting and require human intervention so that the human can then go and figure out what we actually need to do differently.

One thing I just want to call out about this particular example is, you know, so Visa is using a lot of different AI techniques across a lot of different places in our system. These particular techniques are probably a little bit more, you know, a little bit more further along and more developed than some of the most cutting-edge techniques, but they're still -- you know, they’re still on the front end of being applied and serve real production environments.

And one of the reasons that we started with something like a monitoring example, right, where we’re trying to monitor our own performance as opposed to exposing this to consumers, was sometimes when we implement new techniques in a production environment, sort of outside of a laboratory, things don't work exactly the way you expected them to.

And so we wanted to, in this particular case, get a fair amount of experience working with this, some of these cutting-edge techniques, in an environment that was -- that where if they didn't work exactly the way we expected them to, you know, the impact would just be on us, like we would identify a bunch of things we needed to look at that maybe we didn't need to look at as opposed to the impact would be on consumers.

And so, you know, as we talk about these techniques, I think there is enormous promise. You know, I consistently find that models used -- models built using many of these techniques consistently outperform other types of models. But I think it's also important that we develop the practical skills and how do we apply them, how do we understand them, how do we interpret them, how do we make sure that they're doing exactly what we think they're doing as we go forward.

So with that, thank you guys very much. I really appreciate it.

(Applause.)

DR. GOLDMAN: Thank you for that very interesting presentation on how Visa is monitoring for fraud.

Okay. And next we're going to go into some medical uses of artificial intelligence, and we'll begin with Dr. Michael Abramoff, who will look at recent developments in that area.

DR. ABRAMOFF: Anyway, thanks so much for inviting me, having me over. I'm both -- I have a long history in computer science and AI, and it seems that some others also have mentioned that they have been doing this for a while. And you can sort of see my age from the fact that my master's thesis in 1989 was on neural networks to simulate the brain. And so I've been working on this for a while.

I'm also a professor of engineering and also of ophthalmology and I'm a practicing clinician, as well as the founder and CEO of IDx, which is the company that had the first autonomous AI approved by the FDA recently, so it’s actually being used on patients.

And so I want to talk a little bit about the background of why AI in healthcare and specifically in diabetes and specifically in diabetic retinopathy. This is the most important cause of blindness, the most important complication for people with diabetes, that’s what they most fear more than death or amputation, they fear blindness.

And so we know very well what to do about diabetic retinopathy, this complication. I mean, when I see my patients, I know how to treat them, how to operate them, how to manage them. The problem is primarily that we don't find these patients. And so a so-called diabetic eye exam is performed maybe 20 to 30 percent of cases because people don't have symptoms, and so we need to look at the retina, clinicians like me, and that doesn't happen. It’s mostly because it's really hard to get an appointment with me, which is necessary for this to happen.

So the idea is, hey, let’s use AI and imaging to make this better. So this is how it works. I'm not showing a demo, even though it would be only a minute or two, because this is not the appropriate context for that. But it's an autonomous diagnostic AI system. It means it gets a point-of-care result in minutes, but more importantly, there's no human reviewer oversight, so no doctor ever looks at the clinical result. The clinical diagnosis is being made without a physician.

It means that you can now shift specialty diagnostics like what I do as a specialist in an academic hospital to primary care and retail clinics, which, of course, increases the ease for which patients can undergo this exam, and you can also do something about cost of healthcare. Thank you.

It requires, right, a robotic camera because you want to make sure you can do this exam on the vast majority of patients, not just a few. It needs assistive AI for the operator. We will not go into that. And what it requires is a high school graduation for that operator. And it's very important that you need clinical proof that it's safe for patients, right, and we'll go into that in more detail.

And so like I said, I've been doing this for a while and, you know, early on I said, hey, here's an algorithm, in 2000, it can do it, let's just bring this into practice, and that’s, of course, not how it works. You need to do a lot of science, and then you also need to convince the FDA that this is safe, as well as patients and physicians. And I don’t show it on the slide, but my nickname is actually The Retinator, like a terminator, because in 2010 my colleagues were thinking, hey, he’s like a terminator, he will destroy jobs, and he’s also not being safe for patients. And now they think very differently, but it can show you how this fear of AI, you know, is not new. And it’s there and it’s real, and so we also need to manage that.

But, anyway, back to what happened if you do science, and then for many years, you do more science and more science, and you get NIH grants -- thank you -- and NSF grants -- thank you, and many other grants, and then more study sections, but eventually you get to a point where -- we knew that the open source wasn’t going to work, so you need to go through the FDA, raise money to go through the FDA because it took us $22 million to do this, and then eventually you build a company to do all of that.

And so one of the things that happened during that path was that traditionally we use certain features for essentially what we now call AI, and I like the wave of AI so I'm calling it that, but we took a sort of different approach because given the experience in neuroscience, we tried to mimic the brain of clinicians and say, well, clinicians do it this way, why don't we build a computer that does it the same way.

And there's a number of advantages that we now realize that were sort of not even thought about when we did it. And so we built detectors for each of the different visions that you can see in the image of someone with diabetic retinopathy. And on the right, I show this sort of process where the orange images are retinal images, and then you can detect different diseases.

It's like the puppy images and the cookie images that were just shown. We build detectors for the eyes and for the raisins and other aspects of that. And by now, it's being used in clinic. Actually, patients are being diagnosed by the clinicians, but again, no physician oversight.

So there was a scientific stage, I already talked about that, and we learned a lot from that, like the insights from neuroscience and the evolution of mammalian vision story. I cannot read the slides over there, so I have to do it from the big screen.

There were insights from clinical evidence, and it’s really important.

You need to put your AI in a work flow and a clinical work flow, the care pathway, and it needs to fit there, fit with the preferred practice patterns, but the evidence about certain treatments that we already have, and also you need to start thinking about how you actually validate an AI when typically you compare it to humans, but we already know that humans, clinicians like me and my colleagues, have a sensitivity, meaning the ability to detect disease of about 40 percent, so it's pretty low. So we're not really very good at making the difference between subtle degrees of diabetic retinopathy, of this disease.

And so how do you compare an AI to imperfect clinicians, imperfect truth? And it was a big challenge that we needed to solve. And they have insights from interpretation and then poorness of image quality, which is easy to reach in a laboratory setting but very hard to reach in a retail clinic like Walgreens or CVS where there’s no one with any retinal imaging training.

Anyway, I already talked about this approach to essentially basing it on how the visual cortex, the brain of clinicians, solve this problem, and we started to implement that. And that has now a sort of number of advantages that we had not realized at the time but are now pretty logical.

And so -- but before I explain it this way, I want to say that we already did a clinical trial in 2014, where we showed that we did better than clinicians. And we thought, well, that's important. We do better than clinician, that should be enough. And the FDA and we and I agree with them now, they rejected this clinical trial, saying, well, this is not good enough. You need to show it in the actual environment where you want to use it.

So what we did for this clinical trial, it was used in academic ophthalmology clinics where there's experienced photographers, the patient selection is a little bit different, and we showed this result. They said you need to show it in primary care, with the people who already work there, the staff that's already there, which is typically high school graduates and no formal training in any type of retina or retinal imaging.

You need to also decide on the truth, and clinicians are simply not good enough, so how do you compare it, what do you compare it to, and the answer to that was reading centers where it’s very standardized for over 40 years. And you need to do it like I already said with the patients that are already there in those primary care or retail settings, not with a more selective subset of patients.

So that was a clear lesson, and so these are the lessons we and also the FDA, I think, learned from this authorization that we got in April of this year, a lot of things, system validation, all sorts of rules about that. You need the highest level truth so you can compare clinicians and the AI and also say that AI meets certain standards in terms of safety and efficacy.

And also I already talked about the system as a whole. You do not evaluate it just as an AI and reading images; it's actually a system, it’s a robotic camera with the operator, with the patients that are already in primary care. And then you need to preregister a trial, meaning you state what you’re going to analyze, what your hypothesis is, and you try to prove or disprove that hypothesis about safety, efficacy, and what the FDA and we thought was really important, that the vast majority of patients need to be able to undergo a diagnostic result. It's relatively easy to make an AI that does really well on a subset of about 10 percent of patients, but that’s not enough. You need to do it on the vast majority of patients.

I will not talk about this slide. I put these slides together two weeks ago. When I saw the other slides, I realized this is not really the subject of this meeting. This is more regulatory stuff.

But, anyway, so it cleared the path for autonomous AI in general. So it took us a long time to do this but now essentially the rules are set for how you prove autonomous AI making these autonomous decisions. And here are some of the implications already talked about, explainability is now really important.

And there’s a number of advantages that were already discussed, but we actually show that in scientific studies and other groups have now confirmed it. AI avoids racial and ethnic bias because by doing a design this way, we explain it's based on detectors, it’s based on lesions that we already know about for 150 years, clinicians have been using.

When I look at a patient, I look for hemorrhages, for example, and I don't care whether that patient is from Iceland or Kenya, it doesn't matter. If they have the hemorrhage, they have the disease, and the AI does that the same way. But you also avoid the lack of robustness that leads to catastrophic failure. We talked about adversarial images earlier. Well, we look at it as very small perturbations in the images that are not visible to humans that are not visible to an explainable AI, but that CNNs -- typical use of CNNs are very vulnerable to, and we show that essentially you have catastrophic failure in 90 percent or more of cases.

I have two minutes left, right? And like was said already, preregistered clinical trial is really important to prove the safety. It's essentially how we approve drugs, as far as the trial is concerned. And then it needs to fit into the clinic. We already talked about that.

And so I will move to the next slide, which is, well, what are the implications for others following us, and I think it’s very important. It took us a lot of time, but it doesn't mean that others will have the same problem. I think the rules are set now. On the right, you see some implications of doing it the wrong way. I mean, *Bad Blood*, you probably saw the book, and that’s not how we want to do improvements in healthcare and use technology in healthcare.

And one of our competitors had said the following, you know, it doesn't matter if you harm some patients or harm something along the way to improving technology and using technology in, for example, healthcare and this autonomous driving. This appeared in the New York a few weeks ago. So it’s very -- it’s very cogent right now to do this in the right and safe way. So we need to agree on definitions and nomenclature.

You know, technology used in a lab does not directly transfer to what we do in healthcare, and it's very important. Patient safety is very paramount. And if we don't do it right, there will be pushback and we’ll lose all the advantages that AI can have in healthcare for better quality, for better -- you know, lower costs, and for better accessibility, meaning easier for patients to have it.

So, again, I think these are the lessons we learned, that the FDA learned, and I think it will be very important going forward that if you do autonomous AI, we follow these lessons. Thank you.

(Applause.)

DR. GOLDMAN: And thank you, Dr. Abramoff, for that very interesting discussion of how you developed autonomous AI and got FDA approval for your system. Thank you so much.

And now we’re going to have Teresa Zayas Caban, who will continue to look at the use of AI in the medical field.

DR. CABAN: Hi. Good morning, everyone. Very happy to be here and join you to discuss opportunities and considerations of the use of AI in health and healthcare and briefly discuss some activities that my office has engaged in as well as some of our sister agencies in the U.S. Department of Health and Human Services.

A little bit of background before I get started. I work at the Office of the National Coordinator for Health Information Technology. That's a staff division within the Office of the Secretary of the U.S. Department of Health and Human Services. Our charge has been really to facilitate the implementation and adoption of electronic health record systems.

ONC was created by executive order under the Bush Administration and statutorily authorized with the passage of the Recovery Act. There’s a big section in the Recovery Act called the HITECH Act, which created a bunch of different things. One of them you may have heard of. It created an incentive program for eligible hospitals and providers to adopt and meaningfully use an electronic health record system. It also created a certification program which the office I work in runs to certify -- to ensure that an electronic health record system includes certain functionality.

So with that backdrop, the number of electronic health record systems across the U.S. has increased significantly, with about 90-some-odd percent adoption in hospitals and close to that in ambulatory practices. And in 2016 -- in December 2016, the 21st Century Cures Act was passed, which sort of shifted our direction a little bit to focus on now we have these systems in place, how do we make them talk to each other.

So our priorities since then have been to focus on interoperability of electronic health record systems and health IT systems, facilitating the liquidity of health data to enable effective and efficient healthcare delivery as well as reducing provider burden or improving usability of these systems so clinicians have an easier time using them in practice.

So how do we get into AI? Today, I'm going to talk specifically about a report that we released in collaboration with the Agency for Healthcare Research and Quality and the Robert Wood Johnson Foundation that was conducted by an advisory group called JASON. And I'll walk you through the goals of the report and some of the recommendations that came from it.

Leading up to the study, as you may have heard earlier in this panel and earlier today, there's been a lot of progress in AI broadly with the increase in compute and the increase in large data sets that are high quality and well-labeled, a lot of strides have been made in machine learning and artificial intelligence. So with that, we saw also an increase in clinical applications.

And so one of them you may have heard about is in dermatology. And it looks like -- and the most recent one -- my slides are a little changed -- the most recent one is an application developed by Google, really looking at whether an AI application can detect metastatic cancer from a cancer that has not spread. And they've been able to demonstrate this successfully 99 percent of the time. This tool that they’ve developed has actually detected metastatic cancer and distinguished it from a slide that doesn't have cancer.

It was also able to accurately pinpoint the locations of both cancers and observe lesions that, frankly, a pathologist would just not be able to detect with the naked eye. These tools really have the potential to improve care but may require adaptation for successful clinical use. And it is important for them to be deemed effective and be spread across healthcare and different applications, that the technical soundness of their algorithms be tested and demonstrated, that they perform at least as well as the current standard of clinical care. They need to be tested across a wide range of situations and really need to provide improvement, whether that be in patient outcomes, practicality of use, or reduced cost.

I was at the American Medical Informatics Association's annual symposium last week where Jess Mega from Verily Life Sciences gave the opening keynote remarks, and she talked specifically about the need for rigorous testing and appropriate development and application of AI tools for them to be successful and broadly adopted and used in health and healthcare.

Before I go over the goals of the report, I wanted to briefly mention that this is not our first collaboration with JASON. So the Agency for Healthcare Research and Quality and Robert Wood Johnson have previously collaborated on two studies with this group. JASON is an independent group of scientists that have been advising the Executive Branch of the Federal Government for many years. And we specifically engaged them in a study entitled “A Robust Health Data Infrastructure,” which helped inform some of our office's direction in terms of interoperability a few years ago.

We also engaged them in a separate study called “Data for Individual Health,” which looked at how EHRs and health IT could support individual health, allowing individuals to have access to their own health data. And this has actually -- the recommendations from this report have helped spur the health app ecosystem we currently have. A notable example is Apple's use of ONC-recognized standards to implement their health app, which has now enabled individuals to download health data to their iPhones from a whole host of healthcare provider systems.

This third collaboration is the focus of this presentation and began a little over a year ago when we asked JASON to consider how AI could help shape the future of public health, community health, and healthcare delivery. The report focuses on the technical capabilities, limitations, and applications that can be realized in the next ten years.

We asked JASON to consider the opportunities, considerations, and implementation issues around the use of AI in health and healthcare. So under opportunities, there were things -- questions that they asked or looked at where ways where AI may advance the improvement of health and healthcare, evidence that currently exists regarding AI's relevance for health and healthcare, most high value areas, and what kinds of benefits can be defined and assessed.

In terms of considerations, there were three categories that we asked JASON to look at. One was technical considerations; the other one ethical and legal issues; and the last one, workforce issues, which are very important if we're actually going to see increased development of these applications and their implementation across healthcare.

And in implementation, we really asked them to look at other fields and what lessons could be learned that would be relevant to the development and implementation of AI in health and healthcare.

So what did they find? Essentially, JASON concluded that the time may be ripe for the use of AI in health for three reasons that are noted on this slide. Namely, there’s frustration with the existing medical systems, the ubiquity of smart devices, and comfort with at-home services. JASON outlines a series of findings and challenges and makes some recommendations about how to successfully apply AI in health and healthcare.

And I'll go over those quickly, and I have included the link to the report so you can sort of peruse that at your leisure, and I'm happy to answer questions after the session today. So JASON found that overall, AI's beginning to play a growing role in transformative change now underway both in health and healthcare, meaning in and outside of the clinical setting.

So the first challenge they identified was regarding acceptance of AI applications. And so they really recommend supporting work to prepare AI results for rigorous approval procedures, as well as creating testing and validation approaches under conditions that differ from those used for the training set.

With regards to leveraging personal network devices, JASON recommends supporting development of AI applications that can enhance performance of new mobile monitoring devices and apps, developing the necessary data infrastructure to capture the data generated from smart devices to support AI applications and requiring development approaches to ensure privacy and transparency of data use, which is a little bit of what Dr. Kearns alluded to in his remarks earlier this morning.

With regards to the issues around training data sets, they really recommend the development of research, with development and access to research data of labeled and unlabeled health data to support development of AI applications. They suggest that new models are needed to incent the sharing of health data and new paradigms of data ownership.

Some of you may have heard of a movement called Open Science. So there's really an interest in sharing research data sets, but then in healthcare more specifically, there's privacy and security considerations attached to the data. So we need to rethink under what circumstances we can share data to enable both discovery, as well as development of these applications, and validation of these applications so they can be more broadly used.

They also made some recommendations regarding collecting data that are relevant to health but are not systematically collected or integrated into clinical care. So one example is environmental exposure data. But, today, your health is determined mostly by where you live more so than your genome. So we really need to think about what kinds of data are important to health and health care and how we make use of those data and include them into machine learning and AI applications so we make the right kinds of predictions to support whether it be prevention, diagnosis, or treatment.

They really emphasized building on the successes of other domains through competitions, for example, as well as understanding the limitations of AI methods and how they can be applied. They talked about guarding against proliferation of misinformation in this emerging field. As you can imagine, there's a lot of hype about AI generally and specifically in health and health data. So wading through that and ensuring transparency, as well as endorsing best practices by learned bodies.

So since I'm short on time, suffice to say there's a lot of possibilities, there’s emerging applications in health and healthcare, and they range from public health to clinical health, as well as prevention and treatment. Our role is really to work with other agencies to identify what those possibilities are. Our focus is on making data interoperable, to be able to support a development of AI and understanding the data infrastructure issues and what kinds of standards are needed to enable this vision.

And before I close up, I did want to mention two efforts that I thought would be of interest to this audience. So Gina Tourassi heads Health Science Data Institute in the Oak Ridge National Lab that has two big collaborations -- one with the National Cancer Institute and another one with the Veterans Health Administration -- that are really meant to leverage both the compute power and the methodological background that folks at Department of Energy have with the data sources, as well as the research questions and health questions that folks on the other end have to enable new solutions.

With that, I'll stop.

DR. GOLDMAN: Thank you, Teresa. We certainly appreciate your discussion of those issues in the field of medicine.

(Applause.)

DR. GOLDMAN: And it’s a great place to begin the discussion section now. So we've had a lot of discussion of the use of AI in different situations. But at this point, I'd like to put the question squarely on the table. Under what circumstances do our panelists think that it might be better to use artificial intelligence technologies, broadly speaking, rather than traditional algorithms and vice versa? And in considering that, is the selection of the technology generally based on technical considerations or the purpose of the analysis, or are there other practical policy or ethical issues that might add to the decision, some of which we've certainly heard about already today?

So if anybody would like to address that question, please turn your table tent on the side.

So is there anyone -- okay, you would like to? Go ahead, then. Thank you.

MR. RAO: So when we look at when we would use AI versus traditional software programming techniques, the easiest cases for us are anything that -- you need a pattern for -- as we mentioned, we're looking for pattern recognition, so the technical subject matter of what we are trying to do has to be something that we can -- is repeatable and we can train for. So we have to be able to have data that can reveal the problem over and over again so we can train the AI on it. So that's the kind of problem that we can solve with AI. So for us, it has to fit in that category.

If it's a very intuitive decision or a one-off decision or something that’s not going to be repeated, it's not a candidate for us to use AI for, and that's still a candidate for what we refer to is human assistance. So when we think about how to design our software programming, we're looking at what parts can we pull away that are the AI parts and what parts are the parts that are probably always going to be left up to the individual to add their value.

DR. GOLDMAN: Thank you.

DR. KAUTZ: Yeah, so there's a lot of work and interest in human-in-the-loop systems, and that’s probably actually the major category of deployed applications, where we’re not -- it's a person working together with an AI system. I mentioned in my talk examples where people on their own, they simply can't handle the combinatorics of the problem, so that's a good opportunity for using an AI system together with a person.

And I think a number of the people here have talked about these issues of fairness and transparency. There's also some, you know, deep ethical issues. So there has been work, particularly actually in Japan, on robotic friends for the elderly. So these are not truly artificial intelligence systems. They're simulated animals or simulated people that people with diminished capacity might actually come to regard as friends and have an emotional bond to. And I think that could be an example of something we could do but we just should not go down that path.

DR. GOLDMAN: Thank you.

Angela?

MS. GRANGER: Yeah, just to add to, you know, the explainability side is very -- very important, but also the ability to actually implement. If you think about a lot of the techniques that have been talked about, and neural nets, you know, I’ll just pick on because it was mentioned a few times, that's been around for a long time. And in our industry in particular, one of the reasons it hasn't -- it never took off is because the implementation was more difficult.

And so the technology today is there, so when you're doing your research and your analysis, you always have to think about the application and whether or not it can actually be used in production.

MS. MCSHERRY: You know, just to build on what some of the other speakers have said, we are consistently finding when we look at AI techniques -- and I'll compare that what I might think of as more traditional techniques like logistic regression or gradient boosted trees, but when we look at AI techniques, we are consistently finding that those models are outperforming the more traditional techniques.

I think that the -- you know, one of the key challenges is making sure that you have enough data so that the models are not overfit. I think -- I don't know that AI necessarily is inherently more likely to be overfit, but because people are less experienced using it, the human beings are more susceptible to overfitting their models. There are good rules of thumb for how to avoid overfit in something like logistic regression, and the rules of thumb are maybe not as well developed with AI techniques.

I'm pretty optimistic, though, as more people start building these models, those rules of thumb will come as well. So I think, you know, having enough data is one of the key considerations.

And then as Angela said, you need to have enough, you know, computing power, right? So these are computationally expensive models to build, and depending on how you structure them, they can be computationally expensive to run. And as long as you have enough computing power, that's not an issue, but one definitely does need to have enough power.

DR. GOLDMAN: Thank you. That’s very helpful.

DR. ABRAMOFF: Yeah, it’s interesting, I think where you need performance, especially in autonomous AI, you need, you know, techniques that work. And, so for instance, really the techniques that work, and it seems to be that AI is now starting to be essentially whole-vector-based deep learning where you don't know what it's doing.

I don't think that’s what AI is. These deep learning or convolutional neural networks are a technique. There's many different machine-learning techniques that you can all use, and what you saw -- what we do is we combine convolutional neural networks as detectors and there’s sort of a hybrid rule-based system over that and another AI to combine it into an actual dichotomous output.

So there's many different ways, but you still call the entire thing an AI. I think that's valid. And so, for me, it's higher performance, the better you understand it, the better, but AI doesn't necessarily mean that you don't understand it. We showed that we have AI that you can clearly understand exactly what it does.

DR. CABAN: So quickly to build on others' comments, I would say that in healthcare, it's not like there's this set number of circumstances under which AI should be used, but there's certainly some parameters that should be kind of guiding principles that I alluded to during my remarks and that Michael was just alluding to.

You really need to be able to demonstrate that this is as effective or more effective than standard clinical practice. And it really needs to lead to better outcomes. All right? And so if there's enough testing and transparency around whatever AI tool or application is being developed, so long as it's better than the current standard of care and it's been shown to improve something that really needs to be -- that's right for automation.

I really see AI as a tool that can help augment clinical care. Clinicians are extremely busy. There's a lot of data, there's a lot of knowledge that they need to wade through to provide effective care, so think about how AI can help them do that in an unobtrusive manner and in a way that reduces a burden on them to be able to practice.

DR. KEELING: Thank you. So the next question is how accurate are the algorithms in AI tools that we've heard about this morning. And if there is a wide range of accuracy, why is that so? And, also, is the accuracy related to the nature of the tool, the question being asked, or the data being used?

MS. MCSHERRY: So, look, I think, again, in our experience, the AI -- the models that we build with AI, when the competence of the practitioner and the data being made available is the same, and we generally don't suffer from a shortage of data, just given what we do, in those cases, we generally find the AI models to be more accurate. But those two sort of -- when these two things are the same, the data involved and the competence of the practitioner, those are often not actually the same in the real world.

And so I think that the algorithms themselves are -- again, my experience -- very powerful and very effective. And we -- but the models that come out the other side can have a wide range of accuracy because you may or may not have adequate data that’s relevant to the problem being solved and you may or may not have a person who’s building the model who is really effective at structuring that model to get the best possible outcome.

So, you know, when we think about the outputs of these models, there can be a wide range, but my experience has been that has much more to do with the data that’s available and the sort of technical competence of the person building the model than it does the actual algorithms, which again, when we do head-to-head tests seemed to pretty consistently produce outcomes that are better using the advanced AI techniques.

MS. GRANGER: Yeah, and just to add on to that, there's -- you know, credit scoring has been done for many, many years, so it's a very well established predictive use of analytics. And so the lift that you see isn’t -- not probably as great as it is in something that’s more a greenfield that hasn't been done for as long as credit scoring has been.

But when I mentioned earlier in our particular study we saw a 5 percent lift in using some of the more newer techniques outside of regression, what I didn't mention is if you add new data in, you'll also see another 5 percent lift in performance, right? So the data becomes very valuable, regardless of the methodology being used.

DR. ABRAMOFF: It's probably the most challenging problem in medicine, in medical AI, is that what do you compare it to. I and my colleagues differ in about 30 percent of cases. And so if you compare an AI to an individual clinician, when do you know the AI is right and when do you know the clinician is wrong? You will never say that.

And so averaging clinicians will not work much better either. And so we look for ways of doing better. And you can see from our actual trials that we had really good performance -- 97 percent sensitivity catching the disease -- on a data set that was not ultimately to be used in a clinical trial that the FDA authorizes on, where we shot 87 percent sensitivity, the same system. So that risk can be perceived to be very different depending on what you compare it to. And I think it's really, really important that you compare it to the best standard out there, which is usually better than an individual clinician or even a group of clinicians. But that's a challenge that is not really resolved.

DR. GOLDMAN: Okay, so I would like to ask an audience question at this point. I just want to say that we're not going to get to all of the audience questions, but we’re not going to get to all of the moderators’ questions either. And we will hang onto these questions and keep them in the FTC record.

But I'll start with this one. What, if any, efforts do you make to improve your applications of AI after implementation? Do you test for anomalies? Do any third parties review your implementations to provide oversight as you identify problems?

DR. CABAN: So I'll make a general comment, not specific to AI, but like anything else, you have to keep evaluating and testing, so it's part of this continual life cycle, engineering life cycle, whatever you call it in whatever field or discipline you're in. So you have to do that with AI, same as you would with any new tool.

In healthcare in particular, after something is implemented, you need to make sure it’s working as intended and not leading to unintended consequences, undue harm, slower processes, or less effectiveness in care.

DR. ABRAMOFF: Yeah, the FDA required us to build a whole system for continuous efficacy monitoring, meaning we have to consistently monitor that it's up to what we did in the clinical trial.

MS. MCSHERRY: Yeah, I mean, just to pile onto that, I think it's basic good practice that you have to monitor a model. And that’s not -- again, that's not specific to the technique, like you need to do that with any model, whether it’s logistic regression or gradient boosting tree or deep learning or CNN or LSTM or really any algorithm. If you don't monitor the performance of the model, eventually it will degrade and you won't catch it and then you’ll make mistakes.

MS. GRANGER: Yeah, pretty much the same thing I was going to say. Not only that, it's also regulated for us to need to monitor the model and show performance.

MR. RAO: I think in addition to the regular engineering testing, I think for us the new part about AI is understanding that we have to test for inherent bias in the data set, so that was not something that Adobe did traditionally in its software practices. We wrote an algorithm in PhotoShop that was not something we had to think about, but now when we train data to sort out pictures and answer queries and understand content, we actually have an explicit second step of understanding and testing for implicit bias. So that's new because of AI.

DR. GOLDMAN: Thank you.

DR. KEELING: So my question asked, what factors have facilitated the development and advancement of these technologies? Have certain resources and policies facilitated their development?

MS. MCSHERRY: Yeah, look, I think that there are a couple things out there that have been very helpful. First, for us at least, the availability of open source algorithms and the availability of open source data sets has been super helpful. I actually have a person on my team who is a veteran of 20 years of using traditional techniques. And she built her first TensorFlow model a couple months ago, and I said, wow, that's great. And she said, yeah, you can find anything on the internet because, you know, she was able to find, you know, basically everything she needed to go learn this new advanced technique, because it's just all out there.

And so I think the availability, the robust open source environment and the availability of open source tools is something -- has certainly been something that we have benefitted from greatly and we're very supportive of.

DR. KAUTZ: There is also a big advance in hardware around 2007 that made these techniques for deep learning that date back to the '40s and then with additional work done in the ‘80s suddenly scale to real world problems. And this was the discovery by a group of researchers that you could repurpose the graphics processing units of computers that had been developed for computer games and for computer graphics and movies.

These were just the perfect things to use to run neural nets. And they gave a 10,000-fold increase in speed. And you very rarely get a five order of magnitude speed-up. And when that happens, suddenly ideas that could only handle tiny problems, you know, perhaps they could read a zip code, could scale tremendously. So there is that kind of hardware breakthrough.

More recently, companies -- Google, Facebook, Intel, and ARM -- are all coming up with further hardware advances that are tailored for running deep learning systems. And nothing so far will give a 10,000-fold speed-up that’s on the near-term horizon, but perhaps with some radical new ideas about analog circuits, we might see at some point to the next decade another discontinuity in the performance.

MR. RAO: Just on the legal side, what's been helpful, especially for our neural nets, which were trained on images and documents, is in the United States we have fair use exception to the copyright law, and we can use that to allow ourselves to and other communities like us to access publicly available works to train our machine learning.

In contrast, in Europe, they have a copyright directive which currently prohibits that, and it makes it much more difficult to get data to train our neural networks from Europe, and there’s some momentum around changing that, but I do think it's valuable to point out that the legislative framework could also hinder or help development of ML and neural networks.

DR. ABRAMOFF: Yeah, on the regulatory side, I want to do a shout-out to the FDA because they have been extremely understanding and willing to help and make this happen, and now we have the first one approved -- authorized, very careful -- this year. So I think from the regulatory perspective, it's great.

I want to make another remark from the sort of science funding perspective, I’ve been filing for NSF and NIH. That’s also really important starting on, but more importantly, these algorithms existed from Fukushima in the ‘80s. And I used deep learning, you know, back propagation.

I think for us in healthcare, it's always grappling with noisy, insufficient data and sensor design in cameras, et cetera. It that’s what’s really important because I think AI previously failed in medicine, at least, because the inputs were actually noisy. It was usually clinicians hearing patients talk. We then typed it in, and that’s just not good enough to have a really good performance. So the problems we are now having with comparing to clinicians are stemming from the fact that we’re so good and that is because better sensor data is available. A long story but...

DR. CABAN: Yeah, to add to Michael's comment, in healthcare, we struggle with the data quality, data completeness, and missing data. And so that creates a unique set of considerations if these applications or tools are going to be developed using data that’s in electronic health record systems. And there really is a need to better understand what it is we can design with poor data quality and how far we can stretch those models.

DR. GOLDMAN: Well, I really wish that we could continue the discussion, but we are running out of time now. So I would like to ask everyone to join me in thanking our wonderful panel here.

(Applause.)

DR. GOLDMAN: And we'll now have a break for lunch, and we'll be back after that at 1:15.

(End of Panel.)

(Lunch recess.)

PERSPECTIVES ON ETHICS AND COMMON PRINCIPLES IN ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYSIS

MR. TRILLING: Good afternoon everyone. Welcome back from lunch. We are about to resume the hearing. Our next panel will discuss perspectives on ethics and common principles in algorithms, artificial intelligence, and predictive analytics. My name is Jim Trilling. I am an attorney in the FTC's Division of Privacy and Identity Protection, and I will be co-moderating the panel along with Karen Goldman who, if you were tuned in or attending this morning, you have already met. Karen is an attorney in the FTC's Office of Policy Planning.

We are pleased to have a great group of six panelist to discuss ethics and common principles related to artificial intelligence. The format for this panel will be similar to the last one. Each panelists will make a presentation and then we will have a discussion about issues that are raised in the presentations.

We again welcome questions from the audience. Note cards are available for you to provide questions if you want to write them down during the panel.

I am briefly going to introduce our esteemed panelists in the order in which they will be presented. I am sorry, in the order in which they will be presenting. You can find more detailed information about each panelist in the biographies that we have printed and made available on our website.

Our first panelist is James Foulds, an Assistant Professor at the University of Maryland, Baltimore County. Following James will be Mark MacCarthy, Senior Vice President for Public Policy at the Software and Information Industry Association. Then we will hear from Dr. Rumman Chowdhury, the Global Lead for Responsible Artificial Intelligence at Accenture; then Martin Wattenberg, a Senior Research Scientist at Google; then Erika Brown Lee, a Senior Vice President and Assistant General Counsel at Mastercard; and, finally, from Naomi Lefkovitz, a Senior Privacy Policy Advisor at the National Institute of Standards and Technology.

With that, I will turn the microphone over to Professor Foulds.

DR. FOULDS: It is great to be here. This first presentation is on fairness and bias and machine learning and artificial intelligence systems.

So let's make sure we are on the same page. I want to briefly talk about what machine learning is. So we are becoming increasingly aware that machine-learning algorithms, which make predictions based on data, are making a big impact on our lives. A common example that we all deal with is credit scoring, so predicting whether you will repay or default on a loan.

So on the slide, we have a bit of an example of how this might work. So you have some features for every person. So for example, you would have the number of late payments and the amount of credit used, previous defaults, whether or not you are employed, and so on. Then based on these features, you try to make a prediction, in this case, whether you will repay your loan or not. So the features, they are called a feature vector or an instance, and then the thing you are trying to predict is called the class label. So you try to predict the class label Y given the features X.

So these models are trained using a bunch of these feature vectors and they try to imitate what is in the data set, and this is called classification. This is an instance of supervised machine learning.

So it is supervised because the labels are provided.

So there is growing awareness that biases inherent in these kinds of data sets can lead the behavior of machine-learning algorithms to discriminate against certain populations. There are a number of high-profile papers and books on this subject.

So for example, the Executive Office of the previous administration published a report called “Big Data: A Report on Algorithmic Systems, Opportunity and Civil Rights.” And this was really a call to arms to researchers in both computer science and law and other disciplines to start thinking about these problems. So they showed a number of, more or less, hypothetical case studies about how things could go wrong in terms of fairness and bias in machine learning.

This book, “Weapons of Mass Destruction,” by Cathy O’Neil, considers some of the same problem domains, including housing and employment and credit and criminal justice, and goes into greater detail on a number of case studies.

One more book I want to point out is, “Algorithms of Oppression”, by Safiya Noble. So she takes on intersectional feminist approach to understanding this problem of bias and she looked specifically at Google and how Google might, for example, lead to problems of representation. So if you search for the term “black women,” what kind of results do you get compared to if you search for “white women” or “white men.”

So there are also very serious real-world applications where these problems are coming up. There is a system that is already deployed today called COMPAS, the Correctional Offender Management Profiling for Alternative Sanctions. This system is used to predict re-offending in the criminal justice system, and it is being accused of being potentially biased.

So there was an article by ProPublica (Angwin, et al.) in 2016, and they found that this COMPAS system tends to more frequently incorrectly predict that black people will re-offend and end up back in the criminal justice pipeline compared to white people. And it found that the opposite happened for white people, that you were more than twice as likely to be incorrectly predicted that you would not re-offend when you actually did if you were a white person under this system.

So these findings are being disputed, at least Northpointe would like to point out that there are other possible definitions of fairness that this satisfies, but I do not think they dispute the main claims that it does makes these type of errors.

So let’s look at an example to see how this might actually happen. So I am going to show you an example from a blog post by somebody called Rob Speer and the blog post is called, “How to make a racist AI without really trying.” And so he is looking at an application called sentiment analysis. So if you think of reviews such as on Amazon or on Yelp where there is a product or a service and you can type up a review and post it online, we would like to predict whether that review was positive, if you said that was a good product or service or negative, you said that was a bad product or service. So that is a sentiment label we would like to predict positive or negative.

So once again you have feature vectors and you would like to predict the class label. The standard way to do this these days is to use something called a word embedding, which automatically learns for every word in the dictionary a feature vector. And then given those feature vectors for the words, we can try to predict the class label positive or negative.

And so in this blog post, Rob Speer tried to do this and he found that the system, just taking this very standard approach, turned out to be horribly biased. So you can look at the sentiment that the model predicts for stereotypically black names and it finds that the sentiment for those name is, on average, substantially negative, whereas if you look at the sentiment associated with stereotypically white names, then the sentiment is extremely positive. And the sentiment for Arab and Hispanic names is somewhere in between. It is not as high as for white names.

Here is another example. So writers recently reported that Amazon was trying to build an internal tool for recruiting where they would like to predict should we hire this person or not and they found that this system was biased against women. So it seems likely some of the same problems were the cause of these issues that basically whatever was in the data is somewhat discriminatory. For example, if you tried to predict whether you will hire a person or not and then you mostly hired males in the past, then the system is just going to encode that.

So where does this bias come from? So you can look at this article by Barocas and Selbst, “Big

Data’s Disparate Impact.” I will talk through some of the reasons for bias that they point to. So for one, data encodes societal prejudices. So we have already seen an example of sentiment analysis where if you just take data from the internet, let’s say, and people are just saying whatever they want to say, if people are biased and you use that data, you are going to encode those biases.

Data also encodes societal advantages and disadvantages. If certain groups have performed poorly in the past, then the model is just going to learn that.

We also have, by definition, less data for minorities. This could make a classifier less accurate for minority groups. And how you collect the data, this can also be a problem. So if you imagine we only collect data from smartphones, then you only have data on people who have smartphones, so you are going to ignore homeless people, for example, or people who cannot afford a cell phone. This has always been a problem in the past with polling.

If you do a phone poll, then you only find people who have a phone in their home. In the early days of polling that was a problem because it meant that these were the wealthy people, you know, the people who could afford a phone. But, nowadays, most people do not even have a land line and so you are getting a different demographic if you are calling people who have land line phones.

You can also get cases of intentional prejudice. This is sometimes called digital red-lining. To hide that process, this is called masking. There was a case of St. George’s Hospital Medical School -- this was in I think the late ‘70s, early ‘80s when this happened. They encoded what they believed was their own existing process for determining whether they would accept a person into their residency program and they made that system specifically biased against women and minorities. The people making those hiring decisions thought we should not hire women because maybe they are going to get pregnant or leave so we just will not hire them. So they deliberately encoded that into their system.

And so it gets more complicated even if you do not try to deliberately encode prejudice in your system because every variable in your system, all of your features in your feature vectors, are correlated with your protective attributes like gender, race, and age. It affects almost everything else about you. So even if you leave those variables out, then you will, by correlation, still learn some of those same patterns.

So what do we do when we decide to model fairness in an artificial intelligence context? So this is very difficult to do. How do we nail down what is fairness? You know, fairness is -- it is a complicated sociotechnical, political, legal construct, and nobody quite knows what it means. But here are some considerations you might think about. You might want to distinguish between the harms of representation versus harms of outcome.

So when that sentiment analysis system -- a harm of representation is where we see that the system is biased against African Americans. And so in that case, you may be offended by that. Maybe you were upset that that is how you are being represented by the system. But on the other hand, this may actually affect an outcome that happens to you. So if I use those same sentiment classifications or indeed the features that drive them, then I may down weight your CV if you are applying for a job.

Now, there are differences between equality and fairness. So if we try to define fairness as everything is equal for all groups, then we can run into trouble if the groups are actually different.

You have to decide whether to model differences between populations or not, should we treat these as legitimate or should we encode them, and whether to aim to correct biases in society as well as biases in data. So you want to do something like affirmative action.

So a related problem is explainability and transparency. So many of these algorithms are essentially inscrutable black boxes. So it is often very hard to know what these methods are doing. So sometimes there are legal reasons why you have to provide some kind of explanation with these systems, for example, credit scoring in the United States, and then there is the GDPR protections in the European Union.

The law does have some things to say about it other than that. For example, we can just look to Title VII and other anti-discrimination laws, which prohibit employers and other parties from intentional discrimination along lines of gender, race, national origin, and religion.

The basic guidelines for this look at the ratios of probabilities of a positive outcome like hiring a person. And so if I hire all white people, then if I hire black people at less than 80 percent of that rate, then the law says that is an example of discrimination.

The machine-learning community has also tried to deal with these problems. So there has been an explosion of research. It has been going on for at least -- since 2012, but really it has received a lot of attention since around 2016. They have been cropping up new publication venues that are dedicated to fairness and to related issues. There is the FAT/ML Workshop, Fairness, Accountability and Transparency in ML; a spinoff ACM Conference, FAT*; and then there is a AAAI/ACM Conference on AI, Ethics and Society that also has happened in the last two years. In these research communities, there has been a lot of work on defining fairness and algorithms that try to enforce and to measure fairness.

Fairness can also be related to privacy, which is another concern of the FTC. So for example, if I have a system which assigns outcomes to people, like a classifier, it may be possible, based on those classifications, to determine which group you belong to, are you a white male or -- and so on. And if that is the case, then maybe even if our system was fair then somebody could use that to discriminate later on. For example, they could undo the fairness correction that you have carefully done on your system. So this is called the Untrusted Vendor Scenario (Dwork, et al., 2012).

I would also like to point out that fairness should be related to the study of fairness in society, which has long been studied in literature and feminism and especially intersectional feminism. Intersectional feminism makes the argument that systems of oppression built into society lead to systemic disadvantages along intersection dimensions, including gender, race, nationality, sexual orientation, and so on.

So the argument is that if you are a disabled Native American female, you are going to have a very different experience than an able-bodied white male. So, of course, that can be encoded in data and that can lead to problems.

Now, there is a competing notion of fairness called infra-marginality, which just argues that different groups do have different distributions over everything that happens to them, all of their features and so perhaps we should define fairness not as equality, but as the extent to which a system biases above and beyond what is in society.

So in my research, I proposed a definition of fairness which tries to look at both the privacy aspect of fairness and intersectionality and it is also related to fairness in the law, this 80 percent rule where discrimination occurs when more than 80 percent difference between the groups.

So it has privacy and economic guarantees and implements intersectionality and essentially it is an extension of the 80 percent rule. But it allows a sliding scale and it protects multiple protected attributes and provides a privacy interpretation.

So that is it. Here are my contact details if you would like to reach out to me. I have a publicly available pre-print of my work and another pre-print is coming online soon. So thank you.

(Applause.)

DR. MACCARTHY: Hello. My name is Mark MacCarthy. I am hoping that this clicker works.

So I am going to talk a little bit today about some of the principles that my trade association, SIIA, has put together. I want to start off by saying we are not alone in this endeavor. The Belmont Principles, which many of you are familiar with, the principles of respect for persons of beneficence and justice, were developed 30, 40 years ago and they form the basis for the guidelines for human experimentation and the IRB rules that many of you are familiar with from an academic context.

The FAT/ML principles that were just referred to are out there as well. ACM has a new code of professional conduct for their members and for software professionals. And our principles are in the same ballpark. There are two others that I want to mention, both of which have to do with human rights. A group up at the Berkman Center at Harvard has put together a series of very good applications of human rights to some of these ethical principles and to hard cases. And AccessNow has a similar document where they talk about the importance of human rights in the context of AI. So we are not alone in this endeavor.

Our principles are not original. You have probably seen these concepts before. But before I get into them, I want to say a word or two about when to apply these principles because, after all, businesses are engaged in lots of different practices and it may not always be important to think about them from an ethical point of view.

So the way I had sort of set it up is, when the effect of a business policy or procedure has large effects on these values, these principles, then it is important to pay enough attention to do an ethical analysis and that is either positive or negative. If it is a huge infringement of human rights, you have to pay attention to that. If on the other hand your policy or practice increases respect for human rights and provides increased freedom of speech or increased safety or further healthcare, then that is also something that should be taken into consideration. It is not just the negative stuff that you want to pay attention to. So that is one.

The second point is that what is the status of these principles, how should we think about them. And it is a continuum here from the kind of ACM principles, which are really guides to individual behavior, a code of professional responsibility. And then that extends through guides to companies or self-regulatory principles that might be enforced by a group like the Digital Marketing Association and, finally, soft law like the OECD principles that were set up on fair information practices that eventually became law in the European Union in 1985, and then finally law itself.

I think we should think of these principles as guides for company action and not go farther down the continuum. Part of the reason for that is most of these principles are very, very abstract and the key issues are really in the application of these principles, not so much on the articulation of them. And next steps really are not to further refine or provide more detail on these principles. But it is to apply them to particular cases. And that is where we will find all the interesting ethical issues.

So for example, if you want to talk about autonomous cars, the ethical issues involved are much different from the ethical issues involved in autonomous weapons. In the one case, you may need to solve the trolley problem or at least assign responsibility to people when something goes wrong. In the other case, you may not even want to deploy autonomous weapons unless you can figure out who is responsible when a killer robot goes amuck.

So these are very, very different kinds of ways of thinking about it. In other circumstances, the companies disagree about how to apply these kinds of principles. So I do not think they are ready to go beyond just guides for company action at this point. So let's get into it with that as the background.

Human rights. The idea is that when you are engaged in various data practices, collecting data, analyzing data, constructing models, you have to respect internationally recognized principles of human rights, and the sort of ethical thought behind that is your behavior has to really respect the dignity and autonomy of individuals. And you ought to not do that in the abstract, but refer to the documents, the guiding documents that have governed international law for a couple of generations now.

And so which rights are we talking about? Here is a sample from those international instruments, the right to life, privacy, religion, property, freedom of thought, and due process. I think organizations really should be bound to validate those internationally recognized aspects of human rights law.

Justice. Here the real question is distribution. When we start off with a principle that individual people have a right to a fair share of the benefits and burdens of social life and you want to really be in a position where you are not engaged in data practices that disproportionally disadvantage vulnerable groups. In particular, you do not want your data practices to result in applications that are not available to all and are sort of intentionally or even inadvertently restricted based on arbitrary and irrelevant characteristics, which are race, ethnicity, and gender or religion.

The organization should not be totally indifferent to how their goods and services that are produced are distributed. It should be a matter of concern for them who benefits from their new analytical services and products.

But that brings us to the important topic of welfare. The whole goal of creating these new processes and services is to increase human welfare, and to the extent that you can do that through the provision of public services or low cost and high-quality goods and services, you have an ethical obligation to do so.

The last grouping may be a little unfamiliar. It is one of the standard ethical theories. It is called virtue ethics. But the idea is that you want your products and services to contribute in some fashion to human flourishing. This means that you are really trying to help people individually and collectively to be the kind of people who live well together in communities. And many of these concepts are sort of old-fashioned. The words that are used to describe this set of ethical obligations are honesty, courage, moderation, self-control, and the like.

But we all recognize that sometimes business practices can discourage the development of those virtues. All of the attention to things like the addictive nature of some of the internet activities leads you to think that maybe these devices are teaching less in the way of honesty, courage, moderation, and so on, and are more taking advantage of people's weaknesses. So virtues are a very important thing to pay attention to.

In many discussions, these four different perspectives are thought of as sort of alternatives. Pick one. Do you want to do justice or do you want to do rights or do you want to do welfare? Which is it? Our suggestion is try to do them all. Treat them as a kind of checklist and a set of guidelines to go through as you are considering what needs to be done.

But the real issues here -- and this is to repeat a point -- arise in specific domains. And I think it is important to see how these principles are applied in practice because that is where the key ethical issues will really come to the fore.

So to talk about one that was raised before, disparate impact analysis, as was mentioned, a key part of assessing algorithms is to make sure that they comply with the various statutory requirements, including the prohibitions on discrimination. There are three stages of a disparate impact analysis. The first is you have to take a look and see if your algorithms are having a disproportionate adverse impact on people. You have to see if there is a legitimate purpose that is being served by this.

And then the third step is you have to take a look and see if there are alternatives that would have the same effect on your potential purpose without having that disparate impact on vulnerable people.

Three different areas to think about, which groups to assess. The protected classes include race, gender, religion, and ethnicity. One of the things that we encourage our members to think about is expanding to vulnerable groups that are also at risk, but are not explicitly protected by law, and which purposes to assess. The law right now protects eligibility decisions in employment, housing, insurance, and credit.

But there may be other areas that are not covered by existing laws where the decision-making is consequential for people's lives and a company should be thinking about whether or not to have the same kind of disparate impact assessment in those contexts.

So there is a lot more to talk about. I am delighted to be here at this panel. Thank you for having me, and I look forward to the conversation that follows.

(Applause.)

MS. CHOWDHURY: Thank you. My name is Dr. Rumman Chowdhury and I am the Global Lead for Responsible AI at Accenture, and I am going to be talking a bit about understanding algorithmic bias, particularly with a focus on consumer harms.

Much of our narrative today is about primary harms. How do we expand and understand the conversation about secondary harms and what are these secondary consumer harms that we might want to think about?

But, first, as a bit of background into our practice, I have a colleague, Deb Santiago, sitting in the audience today. We lead our responsible AI practice at Accenture. We want to understand the social, regulatory, and economic impact of this technology from development to deployment. We do provide solutions for clients who are very active in the responsible AI community, including groups such as the IEEE, World Economic Forum, World Society of the Arts, et cetera. So we take not only a U.S. perspective, but also a global perspective of industry, government, and citizens.

So just to take a step back and think about why we need ethics. This space is actually very, very new and this panel is very representative of how very new this space is. We have researchers developing research at the same time that practitioners, such as myself, are deploying these solutions to clients. That is pretty rare. So our pipeline needs to be very short, but at the same time, we need to be very, very careful about what we are building and how we are thinking about it.

Most of my time when I first started my job in 2017 was spent building awareness. What is responsible AI? The words we use today we did not even have over a year ago. The way we refer to things, the language that we are using, this evolution of the space to think beyond technological tools to now an evolved conversation about the human rights impact, this is all happening at the pace at which you are seeing it right now.

2018 was a year of action so Accenture was first to market with a fairness tool. We alluded to these concepts of fairness. So my colleagues before me alluded to these concepts of fairness. Our tool is grounded in legal precedents so we have a disparate impact component to our tool, and we specifically think about the impact of the pipeline between the legal and regulatory space to how we are applying this in our solutions.

Finally, what we are thinking about moving ahead is this concept of agency and accountability, which is why I am here today, which is why the FTC is considering artificial intelligence, ethical frameworks, and how it impacts consumers. What we have found from a technical perspective is we cannot solve all the problems and maybe this is obvious to the people in this room, but this is not obvious to Silicon Valley. That we could not solve all the problems by pushing buttons, writing code, and fixing our data.

What we realized, and in the Amazon HR example that Jimmy pointed out is a very good example, that is actually, in my opinion, an example of good governance. They tested a product, they innovated safely, but they actually found that it was an intractable human problem. Their hiring practices were unfair. That is not a data solve. They tried for years to make a data solve. But, ultimately, the question becomes, well, Amazon, now that you have this information, what will you do with it? That is where the systems of agency and accountability come in.

Thinking on a more granular level, if an individual algorithm has a negative outcome, then who is responsible for identifying what that harm is and addressing and reddressing that harm. As citizens and as consumers of this technology, who do I go to if the Amazon recognition system falsely identifies me as a pickpocketer? I know what to do if there is, for example, a biased police officer. We have systems of addressing and reddressing these problems, however we may feel about them. We do not have an infrastructure of addressing and reddressing the harms that are done to people by artificial intelligence.

So to think a bit about what is bias, Jimmy did a really great job of identifying from almost a technologist’s perspective what is bias. We think of bias as a quantifiable value. As a social scientist, I would often call these experimental bias, so things like sentiment analysis, things like imperfect data.

But really the takeaway here is that for us, often when we think of bias, it is a measurable value and often something you can fix if you just throw enough data at it. If you fix your data, you clean your data, you bootstrap your data, we will be able to fix this bias. Or if we change our model, change some parameter, we are endlessly tweaking and changing to address this kind of bias.

However, when nontechnologists talk about data, often we talk about the societal bias. And these four harms that I have listed were developed by the Future of Privacy Forum and I think they encompass the kinds of primary harms that we talk about today, economic loss, loss of opportunity, social detriment, and loss of liberty. Things like the COMPAS algorithm, denying people bail -- I am sorry, denying people parole unfairly. So this is a loss of liberty.

But when we think about bias, we are also often thinking about primary harms. So being specifically denied a job when I am of a protected class is something that is illegal. Now, if we could define all of the harms neatly into those kinds of buckets, frankly, we would not be holding this panel today because existing law would be more than sufficient to address all the harms that are happening or at least the implementation of existing law.

Instead, I want us to think about secondary harms, so this concept I am calling algorithmic determinism. And one thing I want to point at as a good example of algorithm determinism is the filter bubble. Now, what is interesting is we have been talking about the filter bubble for over a decade. We have been living in the filter bubble for more than a decade. The book, “The Filter Bubble”, was published in 2008.

So the question today is, does the filter bubble lead to ideological polarization? And if you are unfamiliar with the concept, a filter bubble is when a recommendation system, an algorithm built by a search engine provider or a media outlet is curating data based on how you are reading information. So what is the incentive of a media company? It is to give you things that you will click on and read. But what happens as a result is ideologically you start to live in an information bubble. You have no idea or concept of what other people are talking about that is different from your notions and your ideas.

Why is this dangerous? The way these algorithms will work often is they will increasingly polarize you towards the opposite end of the people of moving away from the center. And there are two reasons this is dangerous. Number one is the obvious one because I do not know is what is happening in the world and I think that I am always right.

But I think the most dangerous one, number two, is that if someone were to come to me as a human being and say, I actually think a totally different thing from you, I would actually just think they are crazy as in you have no grounding, all the science backs me because that is all I know and all I see, and that inability to communicate on equal ground is really dangerous.

But what I will add to this, this narrative is important because it is not as if we as consumers are battling this, we welcome this. Confirmation bias is a very real thing. We love being right. We love having our opinions affirmed and what happens here is often we are battling our own inner biases. Our desire to be right. We do not like it when we are wrong. We do not like if somebody challenges us. So we are not just battling an algorithm trying to guide us in a particular way; we are also battling our own nature.

So another example -- and this is an example which starts to get into secondary harms, right. There is nothing actually illegal about Netflix targeting users by race. So why are we so upset about it? Why do we think there is a problem with black people being shown images of black people and women being shown, you know, movies with a strong female lead, which is often what I will get in my Netflix queue. But we know that there is something wrong. Otherwise, this would not headlining in *The New York Times*.

And because, as I mentioned, we do not yet have the language in the responsible AI community for many of these things, I invite the term “algorithmic determinism” to think through these secondary harms. Why are we so worried about it? Because we are about a world in which we only identify ourselves by a race, we only identify with people who are of the same race, who are only interested in media that looks exactly like me all the time. What that does is reduce our ability to be empathetic toward other people and other people’s life situation.

So from a quantitative perspective, algorithmic determinism is a measurement bias plus a feedback loop. So a measurement bias ties into what people like myself do which is literally the data bias. And a feedback loop is something -- it is an engineered loop where your output starts to influence your input. If we think about artificial intelligence as an algorithm that learns from its environment, well, if I put something out there and I assume something about the world and then by doing so I make the thing happen and then I use that data to feedback into my algorithm, I am creating a self-reinforcing hypothesis.

So algorithm determinism starts to not only make wrong assumptions -- that is only half of it. The other half is it creates the world in which the wrong assumptions are now true.

So measurement bias, as I mentioned, what you think you are measuring is not what you are actually measuring, and a feedback loop is a structure that causes an output to eventually influence its own input.

So just in conclusion, I invite a conversation around different types of bias. So what does bias mean to different parties as technologists and nontechnologists try to bridge a gap between our lexicon? Let’s make sure we are on the same page about what we mean.

And second is that, as I mentioned, humbly speaking as somebody in the responsible AI community, we are still building our own lexicon, our own language. Our language of harms needs to evolve to embrace algorithmic determinism and the effects of secondary harms. Agencies and bodies like the FTC, who are dedicated to protecting consumers, can also be involved in this conversation and thinking about not just the primary harms, the direct harms to people being denied services, but what are the long-term impacts to society that may happen as a result of algorithmic determinism.

Thank you.

(Applause.)

DR. WATTENBERG: All right. Thank you very much. Thanks to the FTC for having me here. I am delighted this conversation is taking place. And thanks to the other panelists.

So I co-lead a group at Google called the People + AI Research Initiative. Our goal is to make human-AI interaction better, to make it more productive, enjoyable, and fair. We take a broad view of this mission. For one thing, we are interested in all types of people, whether consumers, people who are professionals, like doctors using AI, or engineers or other developers of systems. We think it is important to think about how all of these people work with AI.

We also produce a wide variety of work from fundamental research that we write up and academic publications, educational material, but we also do engineering. We build tools and those tools are the main subject of what I am going to talk about today.

So why are we building tools? Well, let me take a step back and talk a little bit about Google's AI principles. You can see them here. These are principles that sort of guide us internally and externally that we see as a kind of stake in the ground. Some of these, in particular, I think technology can actually help with. You know, we have heard today that technology is not all of the solution, but technology certainly has a role to play in making things better.

In particular, as we seek to avoid bias or avoid reinforcing existing bias, create safe and accountable systems, and just uphold good standards of excellence, tools can be very useful, and I want to talk about a suite of tools that we have released to the open source world. These all have a theme and the theme is helping humans understand AI. For us, we feel the route of -- sort of the best path to moving forward is to increase our knowledge of what is going on with AI systems. You know, it is important I think both from an engineering perspective and to make sure ethically that we are doing the right thing.

You hear a lot that people use the phrase "black box" in talking about machine learning. And it is not wrong in the sense that, you know, it can be difficult to understand certain types of models. The field is moving quickly. However, I think it is inaccurate and there are often many ways that we can actually get a handle on what is going on in systems and then use that knowledge to make improvements.

One very important point I would like to make is that people often talk about transparency as a key value and transparency really has a lot of different meanings here. It is not only as useful to get full knowledge of a system. I mean, just to, you know, give it a kind of silly example of like, you know, if I wave my hand like this, you know, why did I do this. One answer would involve every state of every neuron in my brain, it is not very useful, or the answer might be to make a rhetorical point, which is useful.

Similarly, when you think about AI systems, there are cases where an engineer might need a whole lot of detail to debug a particular issue, but there are cases where a consumer might be overwhelmed by a lot of detail and might need just the type of information they want to make a particular decision or perhaps contest a decision.

Okay. So given that this type of knowledge and understanding of AI systems is important, what can we do to help with that? So one issue is to think about the data that these systems have been trained on. So as we have heard, training data is sort of a key part of any machine-learning model. It really determines the behavior. In fact, arguably, that is the definition of machine learning is that the training data does determine the behavior.

So, in order to understand what a system is doing, it means we need to understand something about the data very often. Now, this is hard because we are dealing often with a lot of data, very complicated data, and, generally speaking, people are not incredibly good at sorting through data unless they have a lot of expert training. Just looking at a huge table of numbers is overwhelming for almost everyone.

But here is a place where technology can help. One approach that my group takes to some of these problems is with data visualization. So one tool that we have released is called “Facets.” And the idea here -- you can see sort of an animation up here that shows this tool in action -- is that it lets you slice and dice this data set in various ways. You can look at quite a lot of data points. You can divide them into groups; you can divide them into subgroups.

One way to look at it using language we have heard today is this is a tool for understanding intersectionality, that we can actually see how different groups interact with each other inside of the data. And often using a tool like this, you can, as a human, start to get a sense of what is going on, what might be driving an issue with your data, what might be potentially an issue that you have not seen yet in behavior. So this is one very important way that we can start to get at what is going on.

Okay. So data is one aspect. What about a model itself? Very often, if you have a machine-learning model that you are trying to analyze, you want to ask it questions. You want to know things like, okay, so I understand how it does on the training data, what if I gave it something that was completely different from anything in the training data set, how would that affect things? Or say it is a classifier and it classifies a data point in a certain direction, you might say, what would change that classification? You might want to fiddle with particular aspects of that data point or ask what is the most similar thing that was classified differently.

So these are natural questions and I think anyone working with machine learning is familiar with this kind of thing. The problem is that they typically require programming, that requires engineering time to do this. That means that stakeholders, people who are not fluent in programming languages may have a harder time getting answers to these questions. So an approach that our group at Google has taken is to create a tool that let's people do this without coding. This is something we call the "What-If Tool" and it is designed exactly to take a machine-learning model in, and then let you pose to it hypothetical questions.

You can see sort of the animation, walking you through a little bit of what is going on there. It is built -- you know, Facets, that visualization we just showed, is part of how this works. And it is kind of a Swiss Army knife for understanding what is going on in a model.

Now, there is something else. In addition to looking at what is happening with an individual data point, we can calculate more global statistics. And this has a lot of helpful uses. One is for thinking about fairness. One thing we can do is if you define particular groups, then you can sort of look at various group-based fairness measures. Now, as we heard earlier, there are actually many different mathematical measures of fairness. I think sorting through these is an important issue for the community.

We do not take a position on this, but we do offer people the option of saying, okay, I would like to measure my system in various ways. We go one step further, then, which is to say, if you have a threshold-based classifier, something very common, then we can do a little optimization and say if it is not fair according to this particular criterion, how would you change the threshold to make it fair or as fair as possible? So this gives you actual actionable feedback that you could use with your system.

Now, again, I want to emphasize that as we have heard so far, fairness is a very deeply complicated sociotechnical issue and in no way do we claim that just tweaking a threshold is going to fix every problem. But it is something that can be an important part of understanding a system and thinking through ways that will lead to a solution.

I want to end with one other technology that our group has developed and this is for looking at neural networks. So 95 percent of the time that you hear people talk about machine-learning systems being black boxes, they are talking about what are called deep neural networks. And the truth is that these networks are complicated. You know, they are typically specified by several very large matrices filled with numbers that can look random at first glance. So they can be difficult to analyze.

They are also often used on data sets that themselves are difficult to understand. A classic example would be image recognition. You know, suppose you have a system that is designed to recognize whether an image is a zebra or not. It is looking at individual pixels and a lot of classical methods will tell you things like, okay, did this particular pixel make a difference to the classification? Did that particular pixel make a difference? It is not super useful looking at individual pixels. Instead, you really want to look at something like, did stripes makes a difference?

So the method that we used is something called TCAV. It stands for “Testing with Concept Activation Vectors.” This is introduced in a recent paper by Been Kim and others. It is released as an open source tool as well. What it does is it uses machine learning to help you understand machine learning. After something is trained, you can give it examples of a concept you are interested in. For example, for stripes, you might give it, you know, say, 20 examples of striped rugs or shirts or whatever. And then you can ask it questions. How sensitive was that zebra classification to the concept of stripes?

And so this is I think a very good example of the type of translucency that is helpful. We are not giving a researcher or a person looking at the network the full matrix of every weight in the neural network, but we are giving them information that is useful at the level that they want in terms of a concept that they are actually interested in.

So I would like to end there, but the point I would like to emphasize is that there are many ways in development we are making real progress in coming up with ways to understand these systems. And I think they no longer need to be considered black boxes.

(Applause.)

MS. LEE: Good afternoon, everyone. My name is Erika Brown Lee, and I am at Mastercard. It is a pleasure to be here, and when I say here I do not just mean Howard University Law School, but participating at the FTC's hearing on competition and consumer protection.

As a former FTC person, I spent ten years at the Commission in roles on the competition side and the consumer protection side. So I appreciate the opportunity to be able to participate in hearings that are covering both sides of the Commission's mission. Say that five times fast.

But before sharing my perspective with you on AI, I thought I would turn back the clock a bit. Not too much, but just for a few years. When you think about -- and some of you in this room might actually be familiar with AI from the concept of a movie that was released sometime ago called "War Games." And when you think about that movie, there was a computer named Joshua who had to actually learn and self-teach so that it would prevent nuclear war.

Well, that movie could have been made credibly in 2018, but it was actually released back in 1982. So, of course, back then, artificial intelligence was a lot more aspirational. But due in part to the computational power -- the increase in computational power you have heard from not only this panel, but earlier in the day -- and access to available data, we now use artificial intelligence as part of our daily lives. And the last panel talked about examples of that, of the innovation behind AI powering healthcare to detailed subway maps to computer vision.

But the agility of AI really presents these opportunities for innovation. And at Mastercard, we use artificial intelligence for fraud protection to make our payment system safer and more secure for cardholders. But as I think you have heard from my colleagues on the panel, there are some opportunities also for some structure around the discussion of ethics in the deployment of AI.

So ethics is somewhat of a diffuse concept just like fairness. It may mean different things to different stakeholders, but several themes have emerged to form a common set of principles. And I wanted to cover a few of those principles today, including transparency, accountability, and privacy by design.

I will start with transparency because of its role in building and maintaining consumer trust, which is a key part of the ethics equation. Consumers need to trust, need to have trust to be able to want to share their data and have confidence in sharing their data with entities. And so openness is a part of the process for gaining and securing and maintaining that trust and it can facilitate that confidence.

But by openness, I am not referring to the publication of algorithms. Martin just talked about the deep neural networks or resource codes. From a consumer perspective, I am not sure how meaningful they would find them. A few months ago, *Harvard Business Review* published an article about a case study involving a Stanford professor, Clifford Nass, who faced a student revolt. What happened? Well, the students in his class claimed that the professor's teaching assistants were grading the same type of material in different ways. And so on their final exams they were getting disparate grades.

It turns out they were right and the professor agreed that there is a disparate outcome, and so as a computer scientist, he designed a technical fix and built a model to adjust the scores. And in the spirit of transparency, he provided by email the full algorithm to the students. But the result was that the students were actually more angry and there were more complaints. So it was hard to reconcile this level of transparency.

So two years after the student protest, some of the professors -- another professor's student decided to do a study to explain what happened. And in that study, the students were provided different levels of transparency about the grades they received on an essay. And it turned out that while medium transparency increased trust significantly, high transparency actually eroded the trust completely.

So the derived conclusion was that users did not necessarily trust black boxes -- you have heard a lot about those -- but that they did not really necessarily need or want full transparency, but actually enough information about the basic insights and the factors driving the decisions that were based on the algorithm.

But context matters. So the idea of transparency varies depending on the context. And so for example, if there is a smart washing machine, the explanation of the decisions behind how to get your clothes clean are quite different in need from decisions about credit scoring or learning or lending, for example. So there is a difference in terms of context.

The other aspect of -- the other principle I want to cover is accountability. And accountability carries forward that level of trust and competence of consumers, but there are several different levels of accountability. On a macro level, accountability can show how AI systems or models are ethically used to create social value. At a more micro level, accountability involves reviewing and assessing those established objectives of an AI system.

And we talked about some of those or you have heard some of those ways in which, from a technical perspective you can accomplish that. But by documenting the review and assessment, it can provide a means of creating that feedback loop that can help in understanding ongoing performance and identify some of those anomalies and unintended -- perhaps unintended consequences that Jimmy was talking about earlier.

Accountability also provides oversight of the technical administrative and administrative controls. We are all familiar with audit, you know, an audit, for example, of access controls. But given the substantial increase of data that is collected by an AI system, those technical controls become even more important.

So the last principle or theme that I wanted to talk about is privacy by design. An important part of the exercise really of using an AI system is to reconcile the tension between the protection of individual privacy and the benefits from pursuing that access to data that I was just talking about that AI needs to be innovative and to work efficiently.

Privacy by design can reconcile those two competing interests. So by imbedding privacy into all of the stages of development -- so from that I mean from design -- well, really from ideation then design, build, testing, deployment, privacy can actually be used as a strategic asset. So for example, the concept in privacy -- one of the key concepts is minimization, which calls for limiting the amount of data that is collected. That may at first seem to be contrary to how AI systems work and what I was just talking about in terms of availability of data.

Well, at a certain point, an AI system may actually not benefit from the increased value or the increased amount of data; in other words, if it is not necessarily improving the success or efficiency of the result. And so limiting data may improve efficiency. Or it may be that data becomes less relevant. And so over time that may also encourage minimization.

Privacy by design we heard a little about that, the legal requirements. Data flows across borders. So even though we are contemplating more of a U.S. perspective here, it is important to consider from a global perspective as well because other jurisdictions have, in fact, restricted, added additional requirements with regard to transparency or consent from the individual to use their data.

And a privacy impact assessment can be used to identify those potential risks and harms to individual privacy and strategies for managing those risks. The idea is that if you incorporate privacy, in particular -- and again it is not sort of a one size fits all, but incorporated appropriately, it can enhance the AI profile.

One other point I wanted to make before concluding is just about data literacy, which is something that goes hand in hand with privacy, and it is part of the broad theme of accountability because data literacy extends from the ideation stage and with the computer scientists and coders all the way through launch of a product.

But I will conclude by saying that as we go forward, it is important to have standards that are consistent, standards that are flexible and inoperable not just in the U.S., but globally, and that ensure meaningful protections of privacy.

So I will stop there and turn it over to Naomi.

(Applause.)

MS. LEFKOVITZ: Okay, thank you. And thank you for having me here today. It is a pleasure.

So I am going to talk a little bit about sort of the research and standard space and also a little bit about where NIST is trying to contribute to some foundational concepts and privacy risk management and engineering and see how they might apply in the AI space.

So at NIST today, we have about -- more than 50 projects that are either contemplated or underway in artificial intelligence and machine learning. And many of these are focused on exploring fundamental questions related to measurement and quantification. And I do not have even barely the time -- I do not have any time, right, in ten minutes to talk about all of these projects. So I really just want to make sort of a key point that you have sort of heard that we have to understand what kind of assurance we can get about the correct operations of AI systems. And I think you have already heard today that even “correct,” right, is sort of a complicated concept and has different view points on that.

But at a bare minimum, right, if we want to have AI systems adhere to ethical frameworks, we really need to understand what that correct operation means in that context. Otherwise, we really do not know if they are going to adhere to them.

So the next set of slides I am going to run through. I am not going to talk to these individually. What I really just want to share with you and I know that these -- I understand these slides will be posted so that you can look at this and get a better sense if you are really interested into where the sort of scope of work is going around various standards.

And so the second point I want to make is that these are not actually finished standards. Nothing that I am going to show you in the next set of slides -- you will see study, you will see all kinds of terms, but none of them are completed standards. This is beginning work.

Why do standards matter? Let me give one example, not in the AI space. So we were working in the identity federation space and wanted to see more privacy-enhancing technologies integrated. And what we quickly discovered was that the underlying protocols on which sort of identity federation is running had never contemplated some of the integration that we wanted to do and literally in terms of sort of like, hey, we want to put this key exchange in here for this privacy-enhancing cryptographic technique and there is no field for that in the protocol. People do not like it when you break protocols, when you break standards because the point is everyone is trying to build their systems to use these standards so that everybody can communicate interoperably.

And so it is actually very important to build in some of these -- what you want out of the system either from ethics or privacy into these standards or be thinking about that because if they get designed, if these sort of underlying standards get designed without that, it is very hard to go back. You can go back and redo the standard, but it is very hard to get your additional technologies sort of retrofitted in there.

And the other point that I want to sort of make is on some of the challenges in this standard space. So you can see that there are these different types of standards. Some of them are very specific, like a standard for ethically-driven nudging for robotic intelligence and autonomous systems. But you see over here in ISO, they have all these different working groups -- that is what WG stands for -- and you can see -- so, for example, SG 1, there is that computational approaches and characteristics of artificial intelligence systems. If you are not thinking about sort of those ethical characteristics, and people in there are not thinking about it, the ones who are actually building that standard, it is not going to done.

So it really takes engagement and you can see there are these multiple groups and they are all working on these different areas. And they do try to have liaisons, but it is challenging and something to be aware of and why NIST encourages everyone who can to get engaged in the standards development so they get developed the way we think they should. So I am going to move on and you can look at these.

Now, I am going to talk a little bit about some of the NIST work. So we introduced some concepts, some constructs around privacy engineering and risk management because we saw some of the same issues that are coming up. What do you do with principles that are sort of this high level and how do you deal with them down at the implementation stage?

And so you know, I will admittedly say that we are using the term “privacy.” But it is an imperfect word, and you will see that I think we cover a lot of the things that people are talking which might, in some people’s minds, go beyond the concept of what they think of as privacy.

The main point here is that first we began to have -- you know, we have some of the same issues like lexicon and language, what are we talking about. Mainly people think that, okay, if I have protected data, I have managed privacy. But, of course, there is more than that. Sometimes we talk about an example with the smart grid, right. So the reasons that some communities were objecting to smart meters was not so much because the utilities could not keep the information secure, but because the smart meters were collecting such detailed information that inferences could be made about their behavior inside their home.

So how do we manage some of those? Well, in security, right, when we want to understand how do we deal with implementation, right, I mean, how do we go from principles and how do we apply them, ee tend to use a security risk model. And so here I think everybody knows there is -- you know, what is the likelihood that a threat can exploit a vulnerability and what is the impact? But how do we apply that in the smart grid space? What is the unauthorized activity that is happening? What is the threat? The smart meter?

So we had some concerns that that was not necessarily the greatest model for the full scope of privacy risks. And so what we said was what is the adverse event and what are some of the things that you have been hearing about? We have heard it in different terms, secondary harms, primary harms. We went with the term "problems" to sort of distinguish from things that might be legally cognizable versus things that are going to be troublesome for people and that organizations may want to manage regardless of whether there is a legal cost to it or not.

So you can see that there is a whole variety of problems. These are nonexhaustive, and you can put sort of anything in there that you want that people can experience. And that allows us to have this model where we can say, what is the likelihood that any kind of processing of data, any particular operation could create some kind of problem for individuals, and what would be the impact? And that is really the heart, right, of where you go from principles to, you know, what people -- my panelists have been talking about which is like, well, how do you change the context? How do you understand how much transparency to have, right?

Well, we can think about sort of the impact and we think about, hey, what do I want this AI to be doing, and how do we want it to impact or not impact individuals? This is where a risk model and risk management processes can come into play.

The final thing I would briefly mention is the other construct that we introduced in our NIST report, is the concept of privacy engineering objectives. And these are essentially additive to the security objectives, confidentiality, integrity, and availability. And so I think you have heard some of the challenges around things like transparency, they can be interpreted very differently. And so, for example, we can elevate that into, as an objective, in terms of what kinds of properties do we want our systems to support, we can say, well, we would like to enable reliable assumptions about processing.

And if we extend that to AI, we could extend that to AI behavior. So we do not necessarily need to know every detail, but we would like to have some reliable assumptions. How much manageability, right, or intervention, right? If I am driving a car, I can make a choice to hit a squirrel or save my child, right. So I can make those choices, and I will take the consequences for that. But what about the AI? Do I have any ability to intervene in whatever programming and decision-making it is making about that?

And then disassociability is really about being able to disassociate information from individuals and devices.

So with that, I will end. Thank you.

(Applause.)

MR. TRILLING: Thank you to each of our panelists for the excellent presentations. To start things off for the discussion portion of the panel, I want to remind our panelists to please turn your name cards to the side if you want to weigh in.

I want to start off with a fairly broad question. So over the course of the day, we have heard references to a number of different ethics concerns and other constructs related to ethics. For example, we have heard about transparency, accountability, privacy, bias, fairness. My question is: Are the ethical concerns raised by artificial intelligence different from the ethical concerns that are raised by traditional computer programming techniques or by human decision-making? And if so, how and why?

James, do you want to start? Jimmy?

DR. FOULDS: Okay. So first, I would say scale is a big difference. Now, so you can build an AI system and then deploy it on millions of people with a few clicks of a button. So just the share scale of potential impact on people, I think that is a big one.

Another one is kind of transparency is different versus human decision-making. In some sense, everything is there in the computer, right? You have a model, or an algorithm that is making decisions and it is all digitally encoded. But it can be difficult to understand what that means or what it is doing.

So Martin was speaking to ways we could try to unpack that, but it is a difficult challenge, whereas as Rumman mentioned if you have a human, you can go and ask them why they made a decision, but we may not be able to do that for algorithms.

MR. TRILLING: Rumman, do you want to go next, please?

MS. CHOWDHURY: Sure. So to echo Jimmy a little bit, I have what I call the three Is, AI is immediate, impactful, and invisible. And what that means is when you deploy an artificial intelligence system, it impacts as wide of an audience base as you have. So you think of a social media company making a change to its algorithm to show you media. It happens right away. There is not, oversimplifying the engineering process here, but there is not like this wait period where you ramp up.

The impact -- and this is what Jimmy was talking about, you touch people’s lives in very meaningful ways with artificial intelligence. And this is different from traditional computer systems and traditional methods of thinking about computation. As opposed to systems like maybe a car or a television, which is tangentially related to our lives, as much as I may love watching Netflix, it is technically tangentially related to my life, the algorithms that influence my life are things that actually are literally impacting my life choices.

And, finally, they are invisible, so this notion of a lack of transparency. But also the fact that I do not always know when there is an algorithm impacting my experience. I am not sure if I am being shown something because it has been hard-coded or selected for me because there is an algorithm. Now, if you think about the notion of bots on social media, those are algorithms posing as human beings. I may think I am being given media or told some information, but I am actually not. It is being curated by an algorithm. So thinking about the difference between AI and traditional computing, specifically with the three Is and importantly about the pervasiveness.

MR. TRILLING: Mark, did you have something to add?

DR. MACCARTHY: Thanks. Let me emphasize the continuity rather than the discontinuities. Many of the same issues that we run across in the older regression analyses models, the credit scores, the recidivism scores that are so controversial right now, provide very good models for how we should think about the ethical issues involved in machine learning and other AI systems.

I think the techniques of explainability, of providing reasons, identifying the major factors that credit scoring companies have been involved in for a generation are useful lessons for AI algorithms as well. You get into a slightly different set of issues when you come to autonomous systems, where the activity really can take place without human intervention. Autonomous weapons where you say, pick your mission and then go execute it, without human intervention, those raise ethical issues that are quite different from standard regression analysis and they deserve different thinking. Same with autonomous cars, to the extent that they are making decisions about what to do on the road without human intervention, those questions really raise some new issues.

But for the most part, in the issues that we deal with on an everyday basis right now, the new systems really are largely similar to the older systems, and many of the principles and many of the techniques for thinking about these problems have been developed for the earlier algorithms and can be applied to the new cases as well.

MR. TRILLING: Martin?

DR. WATTENBERG: Yeah, I just want to add that I think the focus on ethics is actually really beneficial and is helping us even understand existing systems better and what was good about them. So one example that came up earlier is this idea that if you take a human decision-making system and automate it, you might lose the chance for contestability if you do that in a careless way.

And I think what that is telling us is the key issue was the contestability. It is less about automation or not automation and more about what we want as a society around that process. And I think that is an important thing to keep in mind as we think through these issues. Often, we discover thinking about ethics in the context of AI we have clarified our thinking about -- non-AI systems, as well.

DR. GOLDMAN: So I would like to ask a question that is related to the last one in terms of comparing AI to other more traditional methods of analysis. And we have heard a lot of different frameworks and principles for AI, such as the fairness, accountability and transparency, Belmont principles, SIIA, IEEE policy standards. So there are a whole lot of frameworks. And by thinking about these different frameworks and applying them to AI, are we holding them to different standards than would be applied to human or other traditional decision-making?

And, also, perhaps more conflicts and case-by-case question, but how can compliance with these ethical frameworks or principles be measured and by whom?

Maybe we will just go down the line again. James, would you like to start?

DR. FOULDS: So, first, I want to point out that AI systems are engineered, right? They are created. Even though they are run by mysterious algorithms, they are generally put together by a team of humans who work for a company and who will analyze the performance of these systems and measure what they are doing and decide if it is satisfactory. And so to that extent, these systems are actually not that different from other complex systems, such as the creation of automobiles. So my view is that we should hold them to similar standards to other complex engineered systems like creating automobiles or airplanes or spaceships, and so on.

In terms of how to measure these things, so the machine-learning community has put together a large number of definitions of fairness and so on. So these are definitely tools that we could try to use to measure if these methods are fair or not and then we have to probably have a debate about which of them we give the most weight to.

DR. GOLDMAN: Thank you.

Mark?

DR. MACCARTHY: Let me agree with the point that there is a similar set of standards that apply to AI and non-AI systems. I think the principles that I cited are largely usable in many, many different contexts. But that brings me to the measurement question and I do not think there is a good way to measure compliance with principles at that level of abstraction. All of the key issues really are going to be -- wind up being faced when you get to the level of application. And there, I think measurement is the wrong concept because it sounds like if you just add and subtract enough, you will come up with an equation that gives you the right answer.

In fact, these are very, very complicated and difficult ethical question. It is not to say there is no right answer, but it may be the kind of answer that emerges from discussion, debate and reflection on what we want as a society, rather than measuring something and coming up with the right answer.

To go back to the concepts of fairness that were developed before, the computer science community knows perfectly well that they are trying to provide sort of computer science analogs of very basic, legal philosophical and ethical concepts, and they break into two big parts, group fairness versus individual fairness. And people differ in a large part on whether they think fairness is a matter of accuracy and classification and that is it, or they think fairness is a matter of protecting the interests of vulnerable groups, including groups that have been historically disadvantaged.

You get very, very different conceptions of what the discrimination laws are all about, if you take one of those two different points of view, and then you develop very, very different computer measurements of whether you have satisfied those objectives once you bring it down to the level of measurement. But the key concepts are fundamentally ethical, philosophical, and legal. And they are not concepts that are native to computer science.

MS. LEE: Okay, yeah, I think that the question is very interesting because it really poses something that as a community we need to think through, in terms of whether -- you know, how ethics plays out in decisions for AI.

There was a commentary from a German parliamentarian when he was asked about the trolley problem about what the result would be if a trolley is going down -- for those of you who do not know, if a trolley continues straight and does nothing, then it results in the deaths of everyone. But then if it is diverted then, you know, some people die and others do not, so sort of that ethical dilemma. And the response was, well, whether it is a human making that decision or an algorithm making that decision, it is still a tragic result.

So from a human perspective it is just -- it is going to be a split second determination that no one really has time to think about. So you could deploy that almost from a randomness perspective for an algorithm and end up getting the same result. But the creepiness of it comes from that transparency. So how is it -- how is that decision being made? So my panelists have talked about, it comes up a lot more when the impact -- the higher the impact to the individual. And so I do think it flows back to that level of transparency.

But whether it is an AI system or not, levels of transparency and the requirement to provide additional information behind decision-making is long embedded in U.S. law. And so I do not know that it necessarily makes a difference whether it is an AI system or not. To me, it comes down to the impact.

MS. LEFKOVITZ: So I guess I would say that there are sort of different levels of measurement. And part of that has to do with like what are you looking for, right? So I think that has been underlying a lot of the presentations today. And so one reason that we went in the direction of privacy engineering objectives was because of the fair information practice principles are hard to sort of measure. But you can measure what a reliable assumption is, right? You can actually test that.

And so that is one of the reasons why I think the confidentiality, integrity, and availability have been successful as security objectives because they break these things down into pieces that you can then assess. So I think that is part of this conversation today and that we will go on is figuring out what are our objectives and how are we sort of managing risk. What are we looking for? And then we can know what we can measure.

MR. TRILLING: Are there ethical issues that people are raising in relation to artificial intelligence that may be misplaced? And if so, what are some examples?

DR. MACCARTHY: I think the whole notion that artificially intelligent systems will develop consciousness and agency I think is so speculative that it is not a real problem. Yet, is it the kind of thing that absorbs a lot of time and attention, far more than it really deserves, considering that there are real problems associated with these systems that need to be fully addressed.

MR. TRILLING: Rumman?

MS. CHOWDHURY: So I used to start all of my talks by saying there are three things I do not talk about, terminator, hell, and Silicon Valley entrepreneurs saving the world.

(Laughter.)

MS. CHOWDHURY: So I would just add that to the mix.

(Laughter.)

MS. CHOWDHURY: But I would also say that often we over anthropomorphize artificial intelligence. There is -- as humans, we like to impose human features on things. And you think about being a child and your favorite toy, which may have been a bear, but you gave that bear a name and it had a personality, right, or you had an imaginary friend. That is what we, even as adults, we humans like to do.

So one thing that particularly concerns me is a sense of over-responsibility of the algorithm for the negative outputs, a term I call “moral outsourcing,” where by anthropomorphizing the AI and deflecting or pushing all the responsibility on the artificial intelligence, by writing this narrative that it is alive, it is making decisions, et cetera, it has free will, we are removing the responsibility from human beings, and we are scaring ourselves away from the narrative and from the ability to fix these very human problems.

MR. TRILLING: Martin?

DR. WATTENBERG: Yeah, I think echoing what you have heard, I would say it is not possible to over-hype ethics. I think ethics is critical and this focus is really, really good. It may be possible to over-hype AI as we have heard. I think it is a tool. It is an important tool and a very exciting one. But in the end, it is a technology like many others we have dealt with and I think we should deal with it in the same way as we have dealt with other technologies.

DR. GOLDMAN: So this morning in Michael Kearns' presentation, we heard some things about tradeoffs between fairness and accuracy and even tradeoffs between different types of fairness. So I wanted to get this panel's take on those types of tradeoffs and also, what are the considerations that should govern the design of a system in which accuracy and fairness are at issue?

MS. LEE: We clearly all have very strong opinions.

DR. FOULDS: So, yes, there are definitely tradeoffs between accuracy and fairness. Of course, it depends how you define fairness. So there are some definitions of fairness which only consider accuracy as being a good thing. But there are other notions more related to equality or parity where there is a clear tradeoff between fairness and accuracy. So my take on this is an accurate algorithm is not necessarily a fair one because we need to distinguish between the predictive task of classification or making some prediction, assigning an outcome to a person that makes a prediction versus how that is going to be used, which is an economic question, what is the impact of when I used this to make decisions on people's lives, what is it going to do to them? What is the effect on them and on society?

So an example that I like to use is college admissions. So suppose you would like to use a classifier, a machine-learning algorithm to determine whether to admit people to a college. So you could try to predict their GPA.

But we all know that we have a leaky pipeline in STEM and in probably every field and that can be impacted by unfair factors in society. Like if you are poor or marginalized, you are more likely to get sick, you are more likely to have a mental illness, you are more likely to have family members who get sick, you may be far away from healthcare where you live. So you are more likely to have your grade harmed and drop out. So if you just try to predict GPA and use that to determine admissions, then your accurate classifier may not be a fair one.

MS. CHOWDHURY: So the way I think a lot of us are inviting more granularity around the term "fairness," I invite more granularity around the term "accuracy." So this is another one of those examples of technologists and nontechnologists talking past each other. Accuracy means something very, very specific to us. It is a quantifiable value. Again, when we are explaining machine learning -- supervised machine learning -- as having your output, your accuracy is just a measure of how often your testing data was correct.

We take our data. We put it into two piles. We train it on one algorithm and we check our homework on the other. That is our measurement of accuracy. Now, is that a measurement of accuracy we believe in in the real world? Maybe, maybe not. So one might say that sure, minorities underperform. Does that mean that they systematically underperform? That it is the action of being of a particular race that makes you underperform? No, we know that is not true. And this is why we are concerned about proxy variables.

Another thing I am doing additional research in, particularly in algorithm determinism, is this concept of mutability and immutability of variables. Algorithms do not know the difference between things that we can change and things that we cannot change. I cannot change my age; I cannot change my biometrics.

There are things about myself I can change, maybe my educational attainment, my weight, my hair color. But an algorithm does not know the difference between two. So when we think about things like accuracy, how much are we imposing that accuracy as this objective truth or this objective world order, and how is that related to systems of fairness and unfairness in our society?

DR. MACCARTHY: So fairness and accuracy. Let me go back to the Netflix example that you raised earlier. So accuracy, if a company is trying to assess accurately the taste of people in movies, there is a good chance you are going to get racial differences among groups. It turns out people’s tastes differ by race.

Now, should you try to fix this? Is there some unfairness involved in that? Well, you could move away from accuracy towards a kind of group equality. And your reasoning might be, well, you want people to have a diversity of experience, maybe they will see something that is not part of their prior taste and they will learn a little bit more about the way other people live. But the cost might be that there would be a mismatch between the recommendations and people's current taste.

So there is a tradeoff there. People have to think about which one they want as a matter of what we want our society to be like. But it is very similar to what is going on in the recidivism scores. But what this illustrates is that the way we make that tradeoff and the importance that we ascribe to that tradeoff differs by context. In the context of the Netflix example and recommendations for movies, there is one set of considerations.

But in the recidivism situation, there are a whole bunch of different circumstances but a very similar sort of structure. If you assess people's likelihood of re-offending, it is going to turn out that you are going to get racial differences. People re-offend at different rates depending on their group membership.

Now, should you fix this? There are a couple of very strong reasons for thinking that you should. One is that racial bias is endemic in the criminal justice system and it is high time we do something about it. The other is that in the criminal justice system, one of the principles we kind of live by is to protect the innocent. You know, we do not want to catch the guilty so much as protect the innocent. So for both of these reasons you might want to move away from just trying to get as accurate a predictor as you possibly can.

And you can do that by using one of these other concepts of fairness. Group fairness, you can, for example, equalize group error rates. The problem is if you do that, you lose something called predictive parity in the algorithms. And you raise all sorts of complicated legal, philosophical, and ethical questions involving due process, constitutional questions, all of the difficulties about affirmative action are things we have to start to deal with. There is a cost as well in terms of greater risk to public safety by taking that particular direction.

Now, that is where you find the real ethical issues, right. In that kind of tradeoff, you have to talk about it in a concrete context of some particular practice like criminal justice in order to really get your teeth into the ethical problems. It is not going to be solved and we are not going to make process at the level of debating abstract principles. You really have to look at those concrete cases to understand how to make the tradeoffs.

DR. WATTENBERG: I would like to sort of add a kind of practical note to this, which is that I think theoretically you can point to situations where there are real tradeoffs. But practically speaking in my experience, when you have a system, you identify some way that it is unfair and then find a way to fix it. It actually gets better overall. And just to take an example, one of the most common reasons for a system not to be fair is that it has been trained on the wrong data -- data that is not representative of what is happening in the real world that it is being served on. And when you get better data, it is just a blanket improvement or nothing gets worse overall. That is just a good thing.

So in many cases, fairness is just a symptom of other underlying problems, and so I do not think that we should assume there is always a tradeoff between fairness and accuracy.

MS. CHOWDHURY: Sorry to step in, but anecdotally, I have a similar example with our Accenture fairness tool. When we were using a credit risk modeling algorithm to determine whether or not a system was fair or unfair by particular metrics --

disparate impact, predictive parity -- when we actually equalized for predictive parity by gender, we actually found our accuracy rate improved. It improved because we opened up credit opportunities to people who would previously have been denied. So I absolutely agree with you that it is not always a foregone conclusion that fairness and accuracy are a tradeoff.

DR. FOULDS: I have seen a similar situation where overfitting is the problem. So you have a model that is too powerful, that fits too closely to the data, that can harm both accuracy and fairness, and I have seen that happen.

MR. TRILLING: Naomi, did you want to weigh in quickly before we move on to an audience question?

MS. LEFKOVITZ: Yeah, I just wanted to add, I mean, this is why we came up with a privacy risk model, right, because when you are in a tradeoff space, it helps to have a frame of analysis. So in that contextual space, you can understand, well, what is the impact that this measurement of accuracy is having? And how is that impacting or creating problems for individuals? And then can you begin to make decisions and find the solutions that sort of both optimize your accuracy and also minimize those adverse consequences.

MR. TRILLING: One of our audience members has asked, what are the main sources of data that are being used to develop algorithms, and if personal data are a source, how are subjects informed? And I want to relate that to a second audience question, which is if the data are corrupt, is the fault left to data scientists, programmers, or someone else and who is responsible for fixing that?

MS. CHOWDHURY: I think those are both incredibly important questions. So just getting at the concept of data consent, I think there is also an issue here where there is a misunderstanding in the public about what it means to give consent to data and what that relationship with people and data are. So I am going to sort of answer the question, but maybe take the conversation to a little bit different place.

Most people understand a relationship with algorithms and data or data scientists and data to be similar to when you would give your email address to get 10 percent off at some clothing retailer and then they occasionally send you spammy emails. It is a very direct relation. It is purely transactional. And I know the analogy is data is the new oil. But instead I think of data as a new periodic table. Why? Because I can take the same element, hydrogen, and I can use it to make water, something that gives us life, or the hydrogen bomb, right, something that can cause massive amounts of pain and destruction.

And data is very, very similar. What we do not realize is seemingly innocuous data can be used in many different ways. You may not care if a company is picking up the number of steps you walk per day. But when that may influence your insurance premium, you will definitely care.

The problem with getting consent is that we are not even shown what we are giving consent to because the companies which we are giving consent to do not always know how they are going to use them. And, also, are we giving data consent in perpetuity? What if three years from now that is a very viable algorithm where the number of steps I walk per day cross by, you know, other seemingly innocuous pieces of information, plus the IoT from the publicly available cameras that are available in every smart city, will then be used to actually measure my degree of health and, therefore, impact my insurance premium?

What rights -- when I agreed to share my number of steps, that algorithm maybe did not exist. Now that it exists five years later, what rights do I have over it? And these are the kinds of question that we are trying to understand and grapple with and that requires a very fundamental reworking of our relationship as human beings with data.

The other thing I would point out within the consent is we cannot -- even if we take back our information or data or stop sharing, the historical information we have given, we do not have rights over that information. So what must we think about in terms of data we have already provided or we have no control over what we are providing if we are in public, for example?

MR. TRILLING: Erika?

MS. LEE: So I agree with that. I think that the question is such a good one about consent and consumer control over data. It is hard to sort of place and do the chain of activities that can be undertaken once data is ingested. One of the things, as I mentioned earlier, is sort of trying to do a risk assessment. Naomi has talked about this, too, it can be done through a privacy impact assessment and trying to at least identify what the risks are.

One of the mitigation strategies that can partially address the question is sort of anonymization techniques or encryption techniques, but anonymization, in particular, where you are separating the identity of the individual from that data. So to the extent that data can be anonymized, may be a way to use the data -- somebody I think earlier talked about, in addition, differential privacy where you are sort of introducing noise to the data, so it does not affect the integrity and the ability to use the data, but still protects that information.

There are encryption -- also encryption tools like the homomorphic encryption is just an example, but there are strategies that potentially can be deployed to still allow use of the data without sharing or transferring some of that highly personal data.

DR. MACCARTHY: So one last very quick comment. All the difficulties of getting consent that we have been talking about, I think that is one reason why the NIST framework that Naomi was talking about, where the way of thinking is identify a harm that is a possible harm, and then assess the risk of that harm and then take steps to mitigate it, that approach, which puts a lot more of the burden on the data controller than on the individual data subject, may be a very productive way forward.

MR. TRILLING: So the bad news is we are out of time. But the good news is that our next panel, after we have a 15-minute break, I think will be in a good position to pick up the discussion that we have covered on this panel. So please join Karen and me in thanking our panelists for a great discussion.

(Applause.)

MR. TRILLING: And we will return at 3:15.

(End of Panel.)

CONSUMER PROTECTION IMPLICATIONS OF ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYSIS

MS. GEORGE: Good afternoon, everyone. And thank you for sticking around for the last panel of a very full, exciting and informative day. Hopefully, we can keep you engaged through this last panel.

My name is Tiffany George. I am an attorney at the Federal Trade Commission in the Division of Privacy and Identity Protection. With me is my colleague, Katy Worthman, who is an attorney in our Division of Financial Practices, and together we will be co-moderating this panel.

Before I introduce our esteemed speakers, I would like to remind everyone that we have staff in the audience who have comment cards if you have questions. We plan to make this interactive, make it a conversation more than a presentation, and we will welcome your questions and comments throughout the panel and we will take them as they come.

So first, let me introduce our esteemed panelists who have been so gracious to share their time with us today. To my immediate left is Ryan Calo, who is a Professor at the University of Washington School of Law. To his left is Fred Cate, Senior Policy Advisor for the Center for Information

Policy Leadership and a Professor at the Indiana University School of Law. To his left is Jeremy Gillula, who is the Tech Policy Director for the Electronic Frontier Foundation, and to his left is Irene Liu, General Counsel of Checkr. And at my -- at the far end, last but certainly not least, is Marianela Lopez-Galdos, who is the Director of Competition and Regulatory Policy at the Computer and Communications Industry Association. So welcome and thank you.

So throughout the day, we have obviously been talking about algorithms, artificial intelligence, and predictive analytics. And the last panel talked about ethical issues on those topics. In this panel, we would like to drill down even more and talk about the natural outgrowth of those ethical issues, which are the consumer protection implications for AI.

And with that, I would like to open up to the panel to drill down into what are the consumer protection implications?

MR. CALO: Should I start?

MS. GEORGE: Go ahead.

MR. CALO: Okay. Well, thank you very much. I am honored to be here and really admiring of the

Federal Trade Commission’s commitment to keep abreast of emerging technology and a new leadership role in that. One of the innovations of the FTC has been to bring on technical staff very early so that they can actually understand the technologies that they regulate.

So I mean, you know, from a consumer perspective standpoint, there are three I think puzzles that I worry about. And they are each about -- sort of about line drawing I guess you could say. And the first is, does there come a point whereby using machine learning and other techniques of artificial intelligence that companies become -- have such great information and power asymmetry over consumers that we worry about advantage-taking.

So for example, the Federal Trade Commission passed the door-to-door sales rule on the theory that when someone comes to your house you are not in a market context. I mean, this is a sort of much older regulatory innovation, but the idea is that maybe you are home and maybe you are in the middle of cleaning or cooking or something like that and someone comes to your house and tries to sell you something. Well, the door-to-door sales rule is in recognition of the fact that you are not in a consumer position right then.

So what it says is that you have abilities to unravel, for example, the sale and you get certain other things.

Well, what about the fact that increasingly there are objects that are already in our house that are doing the same thing? They are choosing when to approach you. They are leveraging your hard-wired responses to social interactions. Do we need a kind of sales rule for, for example, the Echo? And so that is one sort of -- I do not know how I should speak, but, I mean, that is one example of where you sort of worry about do we need special protections given the intimate position that technologies increasingly have within our worlds.

And then I have a couple of other puzzles, which I will not get into such detail in because we have a lot of people that want to talk, one of which is, are standards of security sufficient? Because the notion of security has been for a long time now the idea that you are hacking into something and you are bypassing a security protocol. But, today, lots of machines can be tricked through a process called adversarial machine learning, the idea being that rather than bypass a security protocol, you just purposely fool the system.

So to talk about Amazon again to keep with the same example, researchers at Georgetown and Berkeley showed that you could play some white noise that none of us would think of as anything other than white noise, but it would surreptitiously cause Amazon to turn on the lights or to purchase something, and so on. It was easily fooled in a way that was problematic.

Our security standards, if you put something out in the world that is easily tricked, a driverless car that can be tricked into perceiving a stop sign as a speed sign very easily, is that an unfairness problem, much like having a system that is not secure?

Then the last thing, and I will stop here, is I really worry quite a bit about the way in which highly intimate information can be derived by what feels to you like very ordinary information, the idea that the intimate can be derived from the available. It begins to break down this notion that somehow there is sensitive information, personal information, and that sharing it is problematic. You know, ultimately, if things about you can you derived from what feels to you like a mundane observation, because of the extremely powerful tools of pattern recognition, you know, perhaps we need to entirely rethink these categories of sensitive and personal and so on.

So I will leave my provocations there for now and pass it along, but thank you for the opportunity to speak.

MR. CATE: Let me add my thanks. It is a pleasure to be here and it is both important and it is terrific that the FTC is doing this. I would say I think we need to sort of start with some maybe more basic principles, not about what the ethical issues are, but rather about the ways in which we raise them.

So one of those we need to recognize that AI is already all around us being used in many ways. And so a lot of today we have talked about AI as if it is coming, as if it is the future, as if somehow we are like ahead of the game in discussing fairness and ethics and issues of consumer harm, whereas once again this is a case where we are behind the curve as we almost always are. It is almost impossible to be ahead of technology. It is being used widely.

Second, I think as with many of the areas involving information and certainly any time we talk about privacy, we are already discovering that people's concerns are highly subjective and contextual. So it really depends if we are talking about my data or your data as to what my concerns are. It depends on what the AI is being used for.

I wear an insulin pump. It uses a very sophisticated AI to try to predict what is happening. I hope it used all the personal data in the world and continues to use all the personal data in the world, but that is because it is keeping me alive. AI that is being used to market to me, I might have very different views about.

And then, third, I would just say I think we will find in that same vein that the types of concern that individuals have may be very different than what society has. So what we know -- I mean, I think about the number of people I know who work in privacy, who spend their days talking about privacy, who really care about privacy, who I know have a half-dozen or more Echo devices at home. So individuals do not always make rational choices and we should recognize we might be concerned about something, but they are voting with their feet and their pocketbook. They know what they are getting into and they are doing it anyway.

Finally, I would say there are the typical set of concerns that we have with almost anything, whether it is a refrigerator or a car, what have you.

Now, it, of course, involves data and that is that it be reliable, that it be accurate to the extent it is something that we care about accuracy. In other words, I want the automatic brakes that use sensors on my car to work consistently, I want them to work only when there is something in front of me, and not just to make it up and start slamming on brakes in the middle of the interstate. And I want to have recourse if they do not work. I want to know where I can go, whether it is a court or the company or an ombudsman or the FTC to get recourse when they do not work.

MR. GILLULA: So we have heard some great things already and I guess we will be jumping into many of these things in more detail. So I will just add two other things. So from a technologist's point of view, I guess I think about two -- I have two other things that I think I would add. One is from just a process perspective in terms of doing consumer protection, it is a lot harder I think to do consumer protection when you do not have visibility into what is going on.

So what I mean by that is AI offers the ability to personalize things to a tremendous degree. I mean, we have already seen this with targeted advertising online. And it is very hard for an outside organization like the FTC to see exactly what ads people are being shown and based on what criteria, unless the company that is actually showing those ads makes a conscious effort, and some have. So to be clear, this is something that is going on, but it is an ongoing problem.

The other -- talking a little bit, Ryan mentioned adversarial examples. The other thing -- and I think we will dive into this a little more, my concern is just unintended errors, problems -- you know, AI is great but it makes decisions in ways that humans do not. So it can make decisions that no human would ever make, you know, even without an adversarial example and that no human would even be thinking that an AI would make. So, you know, if that happens once, you know, it is a one-off, it is an accident.

But then what happens when we are replicating this across all society and we found out that, you know, 1 percent of the time, it will make some decision about a person. And if you talk about the entire population of the U.S., now we are talking about millions and millions of people who are getting a very weird one-off decision. So I think we can talk about that a little more, too. Thanks.

MS. LIU: So as a representative from the industry, from our perspective, there is consumer impact with AI regardless, positive and negative, because there are mistakes that AI makes at times. So the importance from our perspective as part of the company perspective is that we need to make sure that we analyze it up-front. So if you think about privacy, back in the day there was a lot of discussion around privacy by design and companies implementing privacy by design, and how companies did that is they implemented privacy impact assessments in a lot of their products.

Similarly, it is very informative for companies to implement AI by design. In a sense that they should be assessing up-front because AI is out there and we are using it in companies everywhere. So understanding up-front with an impact assessment of all of the different scenarios and how it can impact a consumer in a biased way and in an unbiased way so that you make sure that you understand up-front all of the different scenarios and so that you can weigh the probability and design it in such a way such that fairness plays a role and that AI is not being used to create mistakes or to make unfair decisions.

MS. LOPEZ-GALDOS: Sure. So please let me take one minute to thank you for having me here and also for putting together today's session, which has been very informative.

As my initial remarks, I think one of the things that we have learned throughout the day is that AI is a catch-all term. AI is going to be applied to the credit score system, to the healthcare system, to self-driving cars. So basically it is going to impact all areas of society.

So when discussing and when drilling down what ethical concerns we have and thinking about them from a consumer protection perspective, I would suggest to frame this discussion comparing machine learning to the status quo. And what I mean by this is that maybe we should try to talk about AI in the context of healthcare and try to think whether there is any difference to what we have right now and whether the current regulations focusing on consumer harm or privacy are sufficient to cover the same kind of concerns we have, when machine learning is being used.

And one of the things that we need to acknowledge and -- sorry if I am being a little pessimistic here -- but human beings and human decisions are not perfect either. So we cannot hope to have all decisions made by machines also to be perfect. And some considerations that we might have is that sometimes we might want to deploy AI systems knowingly that they are imperfect because they bring added value to humanity and balancing those tradeoffs I think is going to be key for the future of machine learning and deploying future technology.

MS. WORTHMAN: So in talking about the harms that have come out of -- maybe more specifically in the previous panels, people have spoken about bias, they have spoken about privacy, they have spoken about transparency. In looking at the current FTC enforcement tools, FTC Act, Fair Credit Reporting Act, the Equal Credit Opportunity Act, how well do these statutes address the issues that have been raised by these recent technologies?

And, Irene, I see you nodding, so I am going to start with you.

MS. LIU: Sure. So Checkr -- for those that do not know, Checkr is a background check company that provides a platform to help companies hire faster and in a more compliant fashion. So from our standpoint, we are regulated already by the Fair Credit Reporting Act. So when I think about regulations in AI and the FTC Act in itself, I believe that the FTC Act is drafted broad enough -- Section 5 is so broad in terms of how it says unfair and deceptive practices. So it is used in such a broad way that you could apply any technology to it. So instead of developing technology-specific laws, it is important for regulators to keep in mind that companies like ours and others have other regulations that are not just FTC Act-specific.

So, for example, we have the FCRA Act that requires us to comply with maximum possible accuracy requirements, for example. So if we are producing a report about you as an individual, we need to make sure that it provides maximum possible accuracy. So in doing that, when we are even implementing AI, we need to make sure that AI technology is not making mistakes, it is identifying the right person and that it is creating the accurate report that we need.

So there are a number of other sectors like ours that are governed by different laws. So if you are in healthcare, obviously you have the healthcare FDA laws, and if you are doing robo advisory from a fintech perspective, there are SEC laws. So there are a number of regulations that other companies are also subject to that really put that checks and balances on what companies can do with AI. So I think it is important for regulators to think about that holistically other than just the law that they are regulating.

MR. CATE: So I think this is a great question. I want to take the two laws you mentioned separately. So the Federal Trade Commission Act in Section 5, Unfairness and Deceptive Trade Practices -- actually, I have never met a regulator anywhere in the world who would not like to have that authority because of its breadth, because of the fact it is not limited by a specific type of harm, because of the reach, and, therefore, it applies to new technologies without somebody having to update the law or say, “and we mean artificial intelligence as well.”

Now, having said that, it is kind of end-of-the-road type of law. It does not tell you anything up-front; it does not give you any prospective guidance. These are things the FTC does in other ways and other regulators do in other ways. So I doubt if it is, if you will, going to be adequate to deal with all the challenges that AI might present. But I think it is a very broad flexible law, and in many ways, we give it too little credit for its value in this area.

FCRA I actually think is discovering a new birth, a new life. And again maybe not as exactly as written, this may require some amendment, but this notion of taking something where you use lots of data, that data could be used in ways that affect people, could be used in ways that would not affect people. So you create some general obligations up-front, but you make most of the significant rights, the real actionable rights, depend on something happening, something happening that would trigger an individual's interest in saying, wait a minute, I may have been disadvantaged or harmed -- and then other rights kick in, you know, access to the data or a dispute, a mechanism for dealing with accuracy, and so forth.

I think this could actually be a model that we think of as we identify issues whether it is around AI or big data or other types of intensive data uses, a model for the future as well.

MS. LOPEZ-GALDOS: I think I am going to tend to agree with what Fred and Irene just said. From a European perspective, I think that the U.S. has a technology-agnostic approach to consumer protection, and I do not think that should change with AI because of what I said in the beginning. It is going to affect all aspects of our lives. And what I really think we need to focus on is to see whether potential consumer harms are covered or whether the laws are sufficiently broad to tackle those, and if that happens, then enforce the laws as they are. Some new consumer harms might appear, but I believe that the current system is sufficiently broad to cover those probably. If not, I am sure you will find a way.

But I would not move towards a nontech-agnostic approach. I think that could be bad for innovation and that does not really make much sense if what you are trying to resolve is potential consumer harms. You should focus on whether consumers are being harmed or not when thinking of regulations or not.

MS. LIU: With that said, though, the FTC could definitely play a role in providing guidelines, not necessarily changing laws or creating laws, but the FTC has been known to create guidelines in the past, for example, security in the internet of things, mobile security facial recognition, and those are some of the aspects where the FTC did voice its opinion and provided guidelines to businesses.

Especially in this area of AI where a lot of companies are implementing AI and it is rapidly moving, the FTC could influence in a way by providing a guidance policy statement around their perspective on AI and how to use it fairly and to create a fair system that protects consumers.

MS. GEORGE: So following up on that, obviously, the FTC in 2012 put out our privacy framework and then a couple a years ago we did a report on big data where we sort of laid out how different legal laws that we apply, laws that we enforce could apply in that area. Are there issues that are unique to AI that are not covered by those existing policy statements?

MR. GILLULA: No, go for it, go for it.

MR. CALO: I mean, I think it would be -- I think we need to back up a little bit and say to ourselves, okay, if artificial intelligence is as powerful as proponents say, and if it is going to remake society the way that proponents say, then also we are going to need to have changes to law and legal institutions. In other words, in my view, it is either a bunch of hype or we are going to have to make deep changes to our system. It cannot be like, oh, my God, AI is going to change everything, but nothing should change. That does not actually make a lot of intuitive sense.

But let me just be more concrete. The kinds of harms that I envision with artificial intelligence that may be unique are twofold. There are wrong harms and there are right harms. And the wrong harms are when you get it wrong, and the line-drawing problem that the FTC and others have to figure out is how wrong do you have to get it, how easy is it to get it wrong before there is a problem, you know. And that is true whether it is wildly inaccurate, in which case the credit reporting has something to say about it. But I also think it is just like if something is extremely easy to fool, even though in order to fool the system you do not need to bypass any security protocol, I wonder whether that might constitute unfair design, in much the same way that designing something that is really easy to hack might.

And then there are a set of right harms and these are even harder. These are the kinds of harms that happen when the technology actually is extremely accurate. And we got to ask questions about that, too, right? I mean, so what law, for example, prohibits Uber from using Greyball to figure out whether the people that are in the Uber are law enforcement? You know what I mean? I do not know, but that is an extremely innovative interesting new thing to do is to use algorithms to figure out if maybe the people in the car are going to be police and then avoiding them, right?

And, yet, when the Federal Trade Commission pursued Uber, you all pursued them along a very similar lines to the way in which the FTC pursued Amway decades ago. In other words, the big cardinal sin originally for Uber was that it represented that people were going to make more money on the weekend than they actually were going to make, and that was also Amway's big cardinal sin. But think about the difference between Amway and Uber. I mean, these are -- there is a sea change.

So I think that the Federal Trade Commission Act is quite broad and unfairness and deception is a dream at one level. It has some notice problems as Fred alluded to. But what has to happen is assertiveness. We need to make sure that the Federal Trade Commission has the bandwidth and the mandate to go in there and ask the hard questions, to direct inquiries, and to figure out exactly what is going on. Because I think one of the big problems is is that a lot of the harms that are -- whether they are wrong harms or right harms -- are invisible harms, and they will not come to the fore unless the FTC uses its authority to reach in and find out, or if, you know, some reporter like Julia Angwin figures it out.

So, I mean, I do think we have adequate

tools and I think the FTC is precisely the right agency to do it. But I think they need to be given that mandate to, look, be assertive. This is a new world. That is what we are being told. We are being told this is a new world where everything changes. Well, the FTC should change and it should pursue these things very assertively. That is my own position. I think you all are in the right agency to do it with the right tools. But I think that that assertiveness needs to come back.

MR. GILLULA: So the one thing I would add to that is that transparency can help with that, too. And it may mean that we need some sort of mandated transparency when it comes to AI tools. Now, this is not to say that we would want the same transparency for all AI tools ever. It is going to be an entirely different type of transparency for, you know, how does your washing machine decide the optimal neural network optimized way of washing your clothes versus, you know, how does Uber decide whether or not you should get a ride because it thinks you are law enforcement.

We definitely need some sort of content-specific, but that could help an agency like the FTC be able to see when the sorts of things that Ryan was just talking about are taking place as if we had some sort of mandated transparency.

MR. CATE: I think -- oh, go ahead.

MS. LIU: Go ahead.

MR. CATE: I think another way -- and just to follow on Jeremy’s point. You know, we have always thought of transparency at least in kind of the data or the data privacy world as meaning -- like explaining what you are doing to people who frankly do not care. So we have shoved notices down their throat, we do not read them. We say, oh, we will make them prettier, we will make them shorter, we will make them layered. And at the end of the day, people just do not read notices. That is just the reality. It is a sad, but inconvenient truth.

So one thing we might think about is what would transparency work like in this area. So part of that might be documenting what you are doing. In other words, it might be saying -- building a record in exactly the way we require for human subject research now. So, you know, we have the Belmont principles that led to some law, if you take federal dollars, you have to do this. You then have an IRB, the Institutional Review Board, has to decide when you are going to do things that affect humans. You have to document it. You do not go to an agency to get permission. I mean, the FTC would be overwhelmed if that were the case.

But then if somebody bad happens, if humans are injured, if something unexpected happens, then the institution can be required to produce its documentation that shows it followed a proper procedure. It used the right calculation. Sometimes bad things just happen even if you do everything right.

So I think one of the things we collectively need to be thinking more creatively about is what does transparency look like in a field as rich and fast-moving as AI and big data and other types of high data-intensive fields and what it might be supplemented with, so that we say, you know, maybe it does not mean transparency to the end user who spent all his or her life avoiding transparency, but rather transparency so that it is available for a regulator or for an advocacy group or if it is needed in litigation or for other purposes.

MS. LIU: It is definitely important to have that transparency. And so as companies are building -- again, when I talk about that impact assessment, it is important to think about audit-ability and explainability not only to the consumers, but also potential regulators. And I know Ryan mentioned earlier that AI is huge and it is rapid moving and so potentially the FTC needs a clear authority on that.

From my perspective, if we start that route, we are doing that with everything. I mean, everything was big, mobile was big, internet of things is big. So with every single new technology that emerges to give FTC a clear authority on each one I think is adding burdens and layers of enforcement -- the broad enforcement that they need and that they already have.

So from my perspective, while it is important to have that transparency, enforceability, audit-ability in the companies for AI, in general, I just do not think that we should be creating technology-specific laws or enforcement mechanisms within the FTC for specific technologies because there will be new things that will be rapidly emerging again and we will say this is the next big thing. So at that point, do we build another framework then?

MS. LOPEZ-GALDOS: I was going to react to the discussion taking place right now and say three things. First, I am a big fan of the FTC, so of course they should have the mandate. I think that is the case when consumers are being harmed. And that is respective of whether the harm to consumers is being produced by machine-learning technology or not. I am going to support the technology-agnostic approach to it to be able -- we protect consumers, which is what we care about here.

Then with respect to the tradeoff between accuracy and explainability, which I think is a very, very hard balance to make and a hard analysis to make, I think this is not new. Think about, for example, gender-based price discrimination when it comes to paying for car insurances. Well, people tend to pay -- women tend to pay less than men because basically it is easy to predict based on gender who is going to have more accidents or not. So not everything is new. Some of those tradeoffs and some of the hard analysis we need to make between accuracy of systems and explainability, we are already thinking about them and they already exist in our society.

And the last point I wanted to make is that with respect to transparency, I think it is important, very important, because these systems are very complicated, but I also think we need to have an approach where the different degrees of transparency exist. So for example, if I go to the doctor and what I am trying to find is whether I have breast cancer or not, I do not think I need to know how the machine created all the neural networks to find out that I am going to have breast cancer. I just want to know it is accurate or not and just have a treatment, whereas the doctor might need a different degree of transparency to be able to ascertain the diagnosis.

So I think we need to bring the transparency debate to a more down-to-earth or a more reality-based approach and analyze it on a case-by-case basis.

MR. CALO: I guess -- I mean, first of all, I am not arguing -- personally, I am not arguing that the Federal Trade Commission should get AI authority. It would be kind of cool, you could get little badges with AI division.

(Laughter.)

MR. CALO: That is not what I am arguing.

MR. GILLULA: Would they say “robocop” on them?

MR. CALO: They would say “robocop” on them. This is ingenious.

I mean, I think that what I am saying rather has to do with just how inquisitive the agency is, right? So imagine that we are talking about -- you know, not talking about consumer harms for a moment. We are back now in -- we are talking about people making crystal meth in their houses, you know what I mean? And imagine the way that we regulated that would be we say, listen, take a list of the ingredients that you bought recently and post them in front of your house, and if we walk over them and any of them look like they might be the wrong ingredients, then what we will do is we will follow up or something like that.

No. I mean, there is a hugely different stance when an agency -- a federal agency that has been imbued with enforcement power, is asking pointed, difficult questions, making you explain yourself. There is a big difference between that and a kind of transparency where you just sort of get to pick what you want to share. You know what I mean?

Again, I do not think there should be a special AI task force within the FTC exactly. But rather I think that the FTC needs to use all of its tools and I think that -- you know, listen, frankly, just to speak plainly -- I have tenure now, so I can speak plainly about things.

(Laughter.)

MR. CALO: You know, there has been a history here where the FTC will pursue, more assertively, consumer protection issues and then what happens is Congress or the courts have placed limits on that. So if I were the Federal Trade Commission, I would be constantly thinking about what the right balance is to strike, okay?

But we are in a moment. We have huge companies calling for legislation, okay? We have privacy legislation in California that we are going to want to standardize, and so on. And so this is a big moment, this is a time when we should be expanding is what I am trying to say. But we have the tools and I do not think we need to confer any special authority. I just wanted to add that.

MS. GEORGE: So just a reminder to the audience, if you have questions, please pass in a card. This is a hot bench, so I am sure they would be happy to answer whatever you want to know.

I want to follow up a little bit on the previous discussion. Ryan pointed out that we need to go in and ask the hard questions in order to sort of get to the heart of the matter. So I want to toss it to Jeremy first as to what are the hard questions that we need to answer in order to increase transparency and explainability.

MR. GILLULA: So I was actually -- just as you said that, I was thinking I was going to answer an entirely different question. That is okay. In terms of answering, you know, what are the hard questions about explainability and transparency, I think I agree quite a bit with Fred about -- that transparency to the end user probably is not the right solution. We have seen lots and lots of that and we have seen lots and lots of it fail.

I am actually going to just use my prerogative and answer a slightly different question, which is what are the hard questions that the FTC should be asking not just about explainability and transparency, but about bias and fairness because that is one that I have been thinking about a lot lately.

And I think the right answer there is, if you are talking about a product or a service that has a material impact on someone's life -- and I am going to use that definition pretty broadly; I am even going to include online advertising in that sense -- I think the question you should be asking is what sort of de-biasing or what sort of fairness calibration, what sort of technical measure did you use? What definition of fairness are you using?

Not, you know, we are going to say you must use demographic parity or equality of opportunity or, you know, any of these types -- but we are going to ask which one you picked and did you do the appropriate calibration because if you are not thinking about how you can de-bias the results of your algorithm in some way, then you are really not -- you are clearly not thinking about the problem hard enough. So I would throw that one out there as that is the tough question that the FTC should be asking.

MS. WORTHMAN: Following up a little bit on that, though, is there a risk that the black box of AI is so complicated that you cannot identify what is causing any of the bias?

MR. GILLULA: So it is --

MS. WORTHMAN: Or how to correct it.

MR. GILLULA: So the neat thing about the correction part is there is actually a lot of active research or rather in the last couple of years, some papers published about how to take any black box algorithm and correct it to some level. You know you pick some certain type of fairness metric -- and to be clear, by this, I am talking about a mathematical fairness metric that says we want the same rate of false positives or we want the same rate of false negatives.

As we heard earlier today, there are many, many of these. I think at last count I saw some paper that said there was like two dozen different ones you could choose from. Many of them are incompatible with each other. But you can pick one and you could do it post hoc. You do not need to actually go in and tweak the algorithm. You can do it after the fact to the algorithm.

So I am not too worried about the black box nature or the explainability part of AI. I mean, that was another thing we saw earlier today, too, was -- I think it was a gentleman from Google who was showing how they had done some really neat research on explainability for AI systems, including visualization. So I really do not see the -- for me, the lack of explainability about AI is that companies generally do not want to share information about their algorithms because they are worried that they will lose their secret sauce, and I totally understand that. But it is not about that the algorithms themselves are somehow incomprehensible or unexplainable just because they are on a computer.

MS. WORTHMAN: And this is a question from the audience. Given the decentralized privacy protections in the United States, how will consumers be completely from harm from AI devices where the harm falls outside the regulatory authority of the FTC?

MR. CATE: So I am glad you asked that. First of all, consumers are never going to be completely protected from harm and we should stop talking as if it is possible. And that has always been the case with individual decisions, as well. I know we have had rampant discrimination in individual decisions in credit, in policing, in admissions for decades, for centuries. And so the notion that somehow AI is going to eliminate all that and that is the standard we should hold it to is just setting us up for a fall. I mean, no one will ever -- we will just get rid of AI and we will be the much poorer for it.

Second of all, it is interesting that the question couched this in terms of privacy, a word we have actually not used much up here at all. When we talked about possible harms, privacy was not a prominent one. I mean, we have talked about lots of harms that you might say relate to privacy. But it was interesting, while Jeremy was talking, I think thinking to do the things he is talking about, which are really important, you need data, you need to keep the data. The way you detect, for example, that you are getting a biased result is because you have data revealed the bias.

So we are going to have to recognize that there are going to be some tradeoffs here. In other words, we might say in order to deal with questions of fairness or bias, we actually hang on to more data, or to deal with accuracy, we actually have to hang on to more data. So I think we should at least be honest with each other about the amount of tension between these various goals.

And then just the last thing I would say is the question used the word "harm," which is a word I have used a lot. I like it because nobody knows what it means, so you can comfortably use it. Like Ryan will go write a law review article about it by tonight --

(Laughter.)

MR. GILLULA: -- and show why we are all wrong. But the problem with harm is we do not really know what they are. In other words, it is harm using my data without consent. Is it used in a way that causes me actual injury, physical injury, financial injury, some sort of severe emotional injury? Is it noncompliance with some law relating to data, is that by itself harm? So one thing which I keep saying as we talk about AI, we need to also be talking about what are the things we are trying to maximize and the things we are trying to minimize.

So what do we agree are benefits? That conversation seems to be fairly easy. And what do we agree are the bad things that we would like to minimize? Because they are not going to be consistent. So that is going to be controversial conversation which, frankly, the FTC is in a good position to help lead.

Oh, dear, here it comes.

MR. CALO: No, nothing is coming apart from me decimating -- no, I am just kidding.

(Laughter.)

MR. CALO: No, nothing is coming. First of all, I fundamentally agree with Fred that we have to get over this idea that reflexively just because you gather more information, that is bad. You know what I mean? More information often is very good and it is very good for consumers in many, many, many contexts.

I guess what I would say about harm, I mean, take, for example, a relatively well-known phenomenon, and I believe it was one of the test prep companies, I think it was Princeton Review, was found to be charging more based on zip code for test preparation in Asian American communities, right? That feels like the wrong thing to do and it feels like the kind of thing where I would, if I had a magic wand, go and ask a lot of pointed questions about what other players are doing in this space, what other metrics are using to charge differential prices, and so on. Right?

And the harm, of course, is that because you live in a particular neighborhood and there are certain assumptions about the way that you value test prep, you are paying more money. Sometimes it is not at all hard to see the harm. The harm is just you are paying more money, or with the lifetime value score that *The Wall Street Journal* and later NPR discussed, the idea that you might be on hold for a very long time because you have a low LVS. These items are pretty tangible. They are not well understood, and I want us to be knocking on that door asking lots of questions about these kinds of practices.

MS. LIU: At the same time, AI is not driven by just PII. So there is a lot of data that we are collecting that is anonymized, that is aggregated. And so from a privacy perspective, it may not raise privacy concerns. So it is really important to differentiate those that are creating the -- using information that may not be personally identifiable to you to better your life. And so in that sense, I think it is important for, as we are looking at enforcement mechanisms, to think about privacy, whether it is really impacting the individual, the consumer.

And, secondly, again, you know, I talk about AI by design and also an AI impact framework. And so in that same sense, I really love Google's principles around AI. One of the things that they also emphasize is the importance of privacy by design when you are developing AI frameworks. So that is something that companies should do and I think this is a policy guideline that potentially the FTC can encourage companies to use, just like how it has done before in terms of encouraging privacy by design in AI frameworks as well, too.

MS. LOPEZ-GALDOS: Yeah, so I tend to agree with that, but also when we discuss privacy, we need to understand that privacy means a lot of things to a lot of people and the value of privacy changes on a consumer-by-consumer basis. Like if you ask people whether they care about the environment, on climate change, probably everybody -- almost everybody these days will say, yes, I do care, but then not everybody recycles. So we also need to understand when consumers act rationally or not to discuss the privacy requirement and what degree of privacy we want to protect.

Because I am thinking of -- going back to my previous example, I think everybody would like to be able to use AI to identify potential cancer and to be able to have a more accurate approach that determine whether you are going to be sick or not well in advance, as we saw examples earlier today. I am so sure that people do not want to have their medical records disclosed. And I think that tension is what we need to look into and try to see whether the current laws allow us to ensure that the consumers have their, for example, medical records preserved, which I think we can with the current laws and whether -- how to make sure that society takes advantage of AI, for example, advance the technologies that help us identify potential cancers for all of us.

And I think the discussions need to be, as I said in the beginning, brought to real cases and have honest conversations about what we want and what we do not want because AI can bring a lot of advantages for society and we do not want to stop those. We certainly do want to protect certain privacy elements, for example, medical records, et cetera, but we need to do it on a case-by-case basis and make sure we do not impair the incentives to progress with these technologies for the good of everyone.

MS. GEORGE: So it is interesting that you talk about medical records because I think it was last week I was watching our big data hearing and someone said like most health-related information that is available is not necessarily protected information, it is more commercially available information. And so I am just wondering if AI can apply in that sort of space or how would you design protections around AI in a space where many levels of information are not protected in the traditional sense or where you can infer data from someone from a nonprotected data set?

MR. CATE: So I would argue -- this may be answering a different question -- but that it is not very valuable to be looking at the data; it is much more valuable to be using at the use and its impact on the individual. So it does not matter whether I get your health record or whether I figure it out from the way you use your iPhone, if out of that, I make a conclusion about your health status and I do something with regard to that, presumably the impact on the individual, for example, if it affects insurance rates or it affects willingness of someone to employ you, you know, uses that we would consider suspect, it should not really matter the type of data, it should matter the type of use.

I think AI is going to really drive that home because we can make so many -- remember, AI is all about probabilities and, you know, the probability that that is your face, the probability that that is the way it translates from, you know, Mandarin into English, the probably that whatever, that you have cancer, that you are pregnant, that you have some other condition. And I think we are going to have to stop worrying about where the data -- we may worry about that for other reasons. Maybe there was a promise the data would not be used or there is some contractual issue that has to be dealt with. But rather much more concerned about the use and the impact on the individual.

MS. LIU: Companies should overall just be thinking about what solution they are trying to drive at with AI. So it is important at the design phase, not only thinking about -- like I think a lot of companies when they have data, they think about how can I exploit this. And instead of using that framework, it is important for companies to think about what solution am I trying to solve, what use case am I trying to solve. What user’s life am I trying to make better or easier? And what data can I use from that to help develop a solution or a machine-learning solution that can help better that life of that user.

So with that context, they should also think about what data do they need to collect, so collecting only the data that is needed versus here is a data set that I have, how can I exploit this. That is not necessarily a right framework to go by from a company standpoint, but rather thinking about solution-based, and I think that will help drive solutions that mitigate the consumer harm.

MR. GILLULA: I just want to completely agree with Irene. From an engineering perspective, it is also just bad statistics to say, I have the -- you know, I found some data somewhere, now let me do something with it because how you collect the data is going to influence what data you have, which will influence how accurate it is. And if you are going to do something, if you say, well, you know, I want to use it for some other purpose and so I will just -- you know, I know how to modify the records or I know what portion of the data to throw out, then you already sort of know what conclusion you are trying to get.

I mean, I guess what I am getting at is, for example, say I have some data set that I collected -- never mind. I was going to go into a pretty technical example. If you are curious about that, I am happy to talk with folks afterwards. Let me leave it at that.

(Laughter.)

MS. WORTHMAN: So, Jeremy, one of the things that you mentioned previously was the fact that the lack of -- like the availability of data actually assists in identifying when there has been bias implemented in AI. Could you discuss that just a little bit in a particular instance?

MR. GILLULA: Yeah, so, I mean, so I think -- so what I was talking about was that if you are -- if the purpose of the AI system is to do personalization, so this is not here, now we are not talking about systems that detect if there is breast cancer or like the Adobe presentation that happened earlier where I have some image and I want to find similar stock photos. I am talking more about targeted advertising or making loan decisions, that sort of thing, where the only person who is going to see, generally speaking, the result of some decision is the person that decision applies to and whoever is making the decision.

And so the concern here is that there is just no visibility from the outside world. If I were advertising 30 years ago and I chose to take out an ad in certain magazines, then anyone can go pick up that magazine and look and see what ads am I showing in which magazines and am I showing certain ads to magazines with certain demographics.

Now, it is a lot harder to do that. If I am on Facebook or one of the other various online advertising companies, it is much, much harder. And then they are also doing all sorts of inference to say, who is -- if I want to target people of a certain demographic, with a certain background, with a certain interest, some of that is going to be inferred data. It is not actually going to be data that was actually collected. And so it is even that much harder to be able to tell, you know, am I doing something that is having some unfair impact in some way?

MS. LOPEZ-GALDOS: Yes, I agree, but just a clarification. There is users who decide when they go and select online advertising who they want to target and who they do not want to target. So it is not so much the companies that do. So maybe the bias, we find it in the user we want to target advertise. So you have options. Do you want to target this zip code? Do you want to target this audience? Do you want to target -- there is like a list that you can select. So I think when this cause bias, in that respect, we also need to question ourselves when we make the selections.

MR. GILLULA: Yeah, I mean, part of it does, depending on the particular platform, fall on the platform. So a good example of this is the current complaints against Facebook that their housing and employment ads, the framework was actually designed so that it was easy to discriminate based on race. That was a choice that Facebook made in how they designed their platform and what characteristics they offered in those sorts of advertising. It is totally true that a lot of the time it is -- like is the person who is buying the advertising, it is choices they are making, but also some of it does apply to the platforms and what choices they offer the person who is buying the advertising.

MS. WORTHMAN: Building on that example, you have the Fair Housing Act or you have the Equal Credit Opportunity Act in the credit space where there is -- the FTC has enforced that law in the past. However, taking those in the credit space or housing space sort of out of that, when you have bias what -- this is a question from the audience -- what general authority does the FTC have to attack bias in the Section 5 context? Is it broad enough that it has been used to attack bias on the unfairness authority?

MR. CALO: That is an excellent question. I do not know who you are that asked that very good question. No, I mean, it goes to the issue that Fred and I were talking about, which is the idea of what counts as harm, right? I mean, so especially under the new -- newish, you know, decades old unfairness standard, you have to weigh your regulatory intervention against whether there is actual harm and also you have to look at the benefit to society and to consumers and the market.

So, for example, if you were to bring something that you could show was societally valuable and add a value to the market and to the consumer, but also it had bias in it, even if we were to countenance bias as being a harm, I do not think it would be so obvious that that would constitute a problem, you know. I mean, it is nontrivial.

What I will say is that I am a little surprised that we are not talking a little bit more about deception. In particular, I am a little surprised we are not talking about the way in which a lot of companies have way overclaimed about what this stuff can do. You know what I mean? Way overclaimed. So, I mean, for example, like I was in -- I am not going to name company names, I was going to, but I am not going to.

I was in an airport and I saw this advertisement in the airport and it was just a bunch of people that all looked similar to each other, like it was like a cartoon. And then at the bottom it said, artificial intelligence has already identified who the terrorist is. No, it has not done that. That is incorrect. It has not done that. That is a way overclaim.

So sometimes people -- if you sell nutritional supplements that do not do what they are supposed to do or if you sell anything that does not know what it -- usually you get in trouble for deceiving. But for some reason we are giving these folks that are advertising about AI a pass. I do not understand why, right? I mean, there is verifiable BS out there and I do not understand why it is not deceptive.

MS. GEORGE: I have some more questions from the audience. Can you describe new harms AI may cause? And examples are synthetic video and audio and virtual agents not identifying themselves as virtual.

MR. GILLULA: I can talk a little bit about the virtual agents not identifying themselves as virtual because Electronic Frontier Foundation actually worked on a law in California that was recently passed that was basically an online bot labeling act. And the tricky part of this law, there were bunch of problems with it, we got most of them solved. One is what actually counts as a virtual agent or what counts as a bot.

So let’s restrict ourselves to social media, say. Does it count as too much automation if I am using something -- if I write a bunch of tweets and then schedule them, is that too much automation and I have to disclose that I scheduled them? What about if instead I have a program, because I am a nerd and I wrote up a program that will just automatically generate tweets, but then I review each one and I pick which ones to post? Do I have to disclose that -- do I have to disclose that part? It gets into a very hard line drawing exercise when you are talking about what level of automation.

There are other parts, too, about if you mandate things like an account has to disclose that it is a bot. How do you enforce that? Basically, you have to start unmasking people and then you get into the harms of eliminating anonymous online speech. And anonymous speech is something we value very highly in this country. And if you are starting to eradicate that online, you have to have a pretty good reason.

It looked like Fred was going to say something so I am going to turn it over to him and we will see where this conversation goes.

MR. CATE: I was just going to say I think we are running the risk on this panel of being awfully narrow in what we are thinking about as AI. In other words, it is not just marketing and personalization and targeted tweets. So AI is being used to deliver healthcare. AI is the way we are examining MRI and CT scan images. In other words, the harms we are talking about are not -- a couple of weeks ago, I wrote a letter to the president of a company because I actually still believe presidents of companies love to hear from me, and I got an answer back almost immediately. I sent it electronically. And then I spent the next like three weeks trying to figure out was AI what did that, and I am absolutely convinced that AI is what did it.

Was I harmed by the fact I got a nice response that came from AI rather than the actual president of the company who did not sit around responding to my letter? This just, to me, does not seem like the big issue. On the other hand, not correctly diagnosing melanoma because we are using AI to say is that image likely to be cancerous, that is a harm. That is a really serious harm. Your car not braking for a pedestrian, that is a serious harm that is AI-related. We are using AI in some cities to determine where police are based on calculations of sophisticated data and realtime data about where things are likely to go wrong. So not having police where you actually need them, that is a real harm. People will die because of that harm.

So as seriously as we can take the “I got the wrong ad” or “I got a letter from the CEO that really came from a virtual agent,” I think we need to be opening up our understanding of where AI is being used in this economy, because it is massive. It is being used to where we water crops and do not water crops and it is being used to determine really sophisticated life-changing things. I think it is going to matter to the public frankly more than the email they got. I am not criticizing the email, I care about that. Can I send mine to you and will you tell me did AI generate it?

MR. GILLULA: I can take a look, no

promises.

MR. CATE: Thank you.

MR. CALO: I just want to say that I wrote a paper with a coauthor about the California disclosure requirement that said that it has some First Amendment issues with it. The truth is is that communicating with bots is a new form of communication and one that needs some breathing room. And I think that one potential harm is that these emerging technologies will freak us out and we will overreact. I think that is personally what California did, and I think even the current version, although because of the efforts of the folks at EFF like Jeremy, is much, much better than what it started out as.

I still think and I think my coauthor thinks that it has some First Amendment issues. I mean, you can go check out *Regulating bot speech*. It is coming out in *UCLA Law Review* and see what you think. But I think there are some real harms to overdoing it, too, and I do not mean to be saying we should top down regulate everything.

MS. LOPEZ-GALDOS: Yeah, we have seen some examples of where some jurisdictions willing to regulate up-front AI or the necessary elements for AI to work and that is not necessarily, at this moment at least, the right approach if we want to take advantage of the full potential that machine learning has.

I think what we forget because now we are hearing a lot about AI and machine learning is that, yes, AI has existed for more than 50 years, but really we are only in the nascent moment of the life cycle. We are just beginning to understand the full potential of it. If we start putting barriers to it, we might not be able to allow the engineers to test and see where this can take us.

So I think we need to, yes, worry about consumer harm for sure. And FTC, you know, worry about that and make sure that companies are able to explain their systems and there is no bias, et cetera. But this moment is really the beginning and let us see where we go and let's have more workshops and let's keep learning as we did today and see where the technology stands. Today, this morning, we learned that we talk about the full potential, but what engineers can actually do at this moment is not the full potential of AI. We are still working on the systems and on deep learning, et cetera.

So I think it is very healthy to entertain these discussions. It is extremely important to probably do regular workshops on these matters. But to cross the line and regulate everything, I think it is just too early.

MS. GEORGE: This is open to the whole panel. Are there particular contexts or uses where AI should not be used since it is a nascent area? Should there be a wait-and-see approach in certain instances?

MR. GILLULA: So it is not related to the FTC's domain, but EFF, along with I think like 70 -- maybe 60 or 70 other civil rights organizations and civil liberties organizations, signed a letter saying that AI was not currently appropriate for bail, parole, basically in the criminal justice context, that we do not think the sufficient rules are there and that -- and those are, as Fred was alluding to, seriously life-impacting decisions. And so although it is not in the realm of what the FTC would work on, I think that is one that is important to note where they are starting -- vendors are starting to market and sell AI-related or AI-based risk assessment tools and we definitely do not think it is appropriate.

MS. LOPEZ-GALDOS: I think I agree. As I said earlier, I think the tradeoffs between explainability and accuracy and that tension that exists there is different whereas you apply AI to the potential email you get or whether you are going to incarcerate the person. So I think the debate needs to be done on a very sector-by-sector basis and really take accountability of the realities that that decision is going to encounter. So, for example, if as a result of applying AI, somebody is going to go to jail and we cannot ensure that it is that accurate, I would be more cautious than in other instances, for example.

MS. WORTHMAN: Building off of that a bit, depending on what type of AI is being implemented, what choices and notice should consumers have regarding the use of these types of technologies? I mean, does it vary or should it be constant?

Irene?

MS. LIU: From a notice and consent standpoint, it is definitely important. Most companies also are regulated not only by the FTC Act and others, but especially for those that are doing business in Europe, there is GDPR as well. So there is consent and notice requirement there especially particularly with regards to use of AI. So it is important to provide that notice to comply with GDPR and it is also important to provide that notice for transparency purposes from a consumer standpoint.

But what I liked about Marianela's

perspective earlier is how much transparency you want to give to the users because it could be confusing. So in the example that she provided earlier, the doctor may want to understand what type of database was used versus the patient. So in that context, you do not want to flood consumers with too much information about the type of AI and the database and PII or even any type of information that is being used. It needs to serve its purpose of providing transparency, but not overtransparency that it causes confusion and misleads consumers.

MS. GEORGE: So what does notice and consent look like in an AI context? I will take an example I think that many people can understand, which is credit scoring and credit reports and it is built off of the Fair Credit Reporting Act, which provides for access to a copy of your report, dispute rights, things along those lines. But in that space, you get a report. It lists your credit lines and credit accounts or it lists any criminal history you may have or your educational history. It lists a series of items that you can then look at and see whether or not they are accurate and correct them if they are not. And once those items are corrected, that will have an impact on the ultimate decision. But in AI space how can you --

can you replicate that or what alternatives should there be?

MR. CALO: Fred is a deep expert on notice and choice, one of the leading experts on notice and choice in America. But I will hazard something which is that what is interesting about artificial intelligence, at least when we come to embody it in an agent, which is something that somebody asked about, is that it can be awfully contextual and dynamic.

So I think that we ought to be encouraging -- you know, the possibility of having a conversation with Alexa about Amazon’s privacy practices is, I think, quite exciting, you know, and the idea -- maybe you are anti, but the idea being that you could ask specific questions rather than have some stupid thing that was like really long and you are never going to read it. But you could say, hey, Alexa, can Amazon turn on you remotely to listen in on a conversation, and then get an answer about that. I think that is actually pretty powerful, personally.

MR. GILLULA: I am not anti, I just think maybe only the people on this panel would find it super exciting to have a conversation about Alexa about --

(Laughter.)

MR. GILLULA: Which is not to say it would be me, I agree. I just do not think the vast majority of consumers would get a ton out of it.

MR. CALO: I mean, I think it is critical when you are thinking about emerging technology generally not just think focus on what is loss, but what new affordances might be there or what you might gain. I think that these things are quite powerful. I think we are getting to a place where natural conversations are becoming more viable and I think that we should therefore -- I mean, if you think about it, notice and choice, we have been operating under basically Gutenberg technology all this time, right?

We just publish a long thing whether it is a digital or a print, just a bunch of words on a page. Yet, you know, here we have companies that are doing these amazing things about organizing information and gauging you and so on. Anyway, I think there is a lot of innovation that could be occurring with notice. And part of it would be to contextualize and actually answer questions about this consumer instead of just having something that no one reads.

MR. CATE: I would echo everything Ryan said. I would just like to make two points. One is we put in the record a paper that I did with some colleagues at the Center for Information Policy Leadership about AI, how it is used today and some of the issues it raises, and one of the things we talk about in there is the way AI is already being used to enhance privacy protections, not just to make them more easily understood or explainable, but to actually activate them. So in other words, you can identify somebody's privacy preferences as they start expressing them and then you can start predicting what they will be so that you offer them the default they are more likely to care about. Rather than the default that you want, you try to give them the default that they want.

I would say just, in general, though, back to the original question on notice and choice. As I said earlier, we have relied on this largely because we have not known what else to rely on for 50 years now, with not a lot to show for it. And so I think we should recognize that notice should be used and choice only where there is something meaningful to tell the individual and only where there is something they can do about it. So I think it is terrific when my iPhone says, did you know this app is using your contacts, do you want to permit that? That is meaningful notice and I can do something. I can say yes or no, I can alter it.

But making my doctor add another paragraph to the 65 paragraphs of the HIPAA notice saying, by the way, your scans are going to be read by AI and, by the way, you have no choice about that whatsoever because it is far more accurate than humans, I am not sure that is overly meaningful. I think we have to be very contextual with notice because the effect when we do not is that we just teach people to ignore all of it. We get people in the habit of knowing that notice is meaningless and so they do not read it, whereas if we would use notice when there actually is something worth telling them and something they can do about it, we might resurrect notice as a meaningful data protection tool.

Now, having said that, I am not disagreeing with Irene. The law requires, both in Europe and in some industries in the United States, notice and choice, it is just bad law. In other words, it is causing people to ignore these notices by providing them when you cannot do anything about them and nobody cares.

So one of the things we often talk about at universities is, you know, a teachable moment. You know, you can only teach someone when they are interested in learning. Similarly, you can only give meaningful notice when there is something that is going to cause them to care about it. And that cannot be they woke up in the morning or they went to a doctor's office. It might very well be where there is an event, there is something happening, there has been some effect on them, there is some reason that they would care, and then using the tools that Ryan was talking about would be fabulous to really make notice meaningful and interactive.

MS. LIU: There is always a conflict within companies with product design when you are trying to design products that is easy to use and that is easy to understand. When you are throwing in all sorts of consents and notices, it can make it really difficult. And so there is often a conflict between the lawyers and the product design teams about how can we make it look beautiful without all your verbiage. So that is something that we struggle with.

And I completely agree with Fred that meaningful consent is ultimately more beneficial to society and to consumers for how their information is being used and how the company is using it versus just providing our lengthy privacy policies that most companies have.

MS. GEORGE: And as a corollary to that, does the notion of opt-out work in an AI context and does that vary based on I think the stage of the product life cycle, be that data collection, you know, product design when it is rolled out to market and being used or other instances?

MS. LIU: Jeremy and I were talking about this earlier. So from a GDPR standpoint, you do have a right to erase your data. So there is an obligation for companies to be able to remove that data. And depending on how you configure that information, it can be difficult. So that is something that you need to think about from the beginning in the design phase to ensure that companies, especially with the California Privacy Act as well, it is important to design these products in such a way that there is an opt-out notion.

To opt out of AI, typically if a company -- if someone wants to opt out of AI completely, that is like let's say if I am using Netflix and I want to opt out of using the choices, the different types of videos or shows that they are showing to me, it is basically opting out of using Netflix completely. So you have to think about, like, are you trying to opt out of the product or are you trying to opt out of the database use as well? So there are different ways of viewing opt out, and I think Jeremy can probably talk more about the technological ways of opting out.

MR. GILLULA: Yeah, there has been some recent papers that show that for neural networks you can actually reconstruct what the training data was if you are given sufficient time and access to -- and able to run test data through the neural network, which basically means that if I am a service and I used your data to train my neural network, I cannot remove your -- the fact that you are -- the imprint your data has left on my neural network basically without retraining it from scratch and retraining it, once again, without your data. So it is technically -- is it technically possible? Yes. Is that potentially a huge burden on the company? Potentially, yes.

Then there is the other question of, how much benefit do you get from having your imprint removed from whatever model was generated? Because it does take quite a bit of effort to reconstruct all of the training data, and so that is in the unlikely but feasible attack. So there we do have to get sort of into this balancing act a little bit.

MS. WORTHMAN: Another question from the

audience. In cases where autonomous systems result in consumer harm, who should be held liable and to what degree?

MR. GILLULA: Just send the robots to jail.

(Laughter.)

MR. CALO: Well, I mean, I think that is a genuine puzzle. I mean, so you have -- in criminal law and in tort law, we generally require that you do something either on purpose or that you -- a reasonable person would be able to foresee the category of harm that occurred, right? And so when you, for example, have a bot, which this really happened, that is supposed to buy things randomly on the web and buys methamphetamine and the police come and say, you know, you bought methamphetamine, and you say, no, no, it was the bot, right?

Or in another instance, where a company made a bot that was arguably hacked into, but at least was subverted by trolls that wound up denying the holocaust which is not lawful in some jurisdictions where -- that had access to this bot. You know, you would be sort of hard pressed to bring a criminal case to it. And certainly in many categories where -- something happens where the system just behaves in a way that was not anticipated, you do not have what is called proximate causation for purpose of bringing a tort lawsuit, which is what I teach.

And that is not a great place to be because you wind up in a situation where you have victims, but not perpetrators. And I do not know how much that would really matter to FTC enforcement, specifically, because I think you could get around it just by saying, look, you created these conditions that were deceptive or unreasonable and these unexpected things happened, but something was going to go wrong. But I think it is pretty serious in tort and criminal law. I think it is hard.

MS. WORTHMAN: We have also had a question from the audience about retail price discrimination at the individual consumer level and what is the material harm to the consumer in price discrimination and maybe price discrimination can be sort of whether or not it is advertising different things, not on a prohibited basis under ECOA, but just because you are using a different type of computer, because you are purchasing tickets on your mobile rather than on a laptop. What is the harm, what is the cost-benefit analysis in that particular instance?

MR. CATE: So this is a place where actually notice would be quite useful. This would be much more, in other words, to say if you visit this website, we are going to use pricing based on information about where you are coming from, the computer you are using, whatever because it would then empower you to say, well, I am going to go have my friend check and see what the price is to see if I can get a better price. In other words, that would be actionable notice, you could really do it. And by the way, having to disclose it would probably slow people down -- companies down actually wanting to do that.

I mean, remember, we have discriminated on price for forever, I mean, for generations. Every time you fly, there could not be -- there is more discrimination for all sorts of reasons, how long you are willing to stay, what nights you will stay, and so forth. We discriminate based on zip code, we discriminate based on all sorts of other information that have been imperfect. Now, we are going to be able to discriminate better. I mean, we are going to have both better technology and better data and the two together are going to make much more precise discrimination. You know exactly what I will pay. eBay knows exactly what I will pay for something because it has watched me pay that for years.

So this is actually a place where you could say, first of all, we need to figure out is that a harm. Is it something we are going to say is unfair? Is it something that we are going to say causes injury? And if not, maybe disclosure is sufficient. To say, look, we are not willing to say we are going to prohibit it, but we are going to say you get notice. So now, you can figure out if you want to try to come back at the system the other way. They are doing it to you, can you do it to them?

But this is why we have to remember, again, it is going to be very contextual and it is not something new. It is not something AI is going to create. AI is going to make it better in the sense of potentially more precise or more tailored.

MR. CALO: I will give you my two favorite examples of price discrimination after -- I mean, and by favorite, I do not mean I like them. One of them was a couple of years ago a marketing firm was using this tool to figure out when women felt worse about themselves and they labeled these “prime vulnerability moments.” And they suggested that perhaps you should advertise or charge people more during those moments, you know what I mean? That strikes me as not a very good use of price discrimination.

Another one of my favorites, although they claim they never did this, is when Uber experimented with figuring out whether you would be more willing to pay surge prices when your battery was low on your phone because maybe you would get stranded there. Lovely, also. They say they have never done this and I believe them about that.

The issue is not price discrimination. The issue is taking advantage of people, which happens, it happens a lot. And, yes, from an economic perspective, better information is better. Maybe we would worry at one level about all the social surplus going to the firm. You know, they know your reservation price. There is no windfall for you because they charge you more if you would be willing to pay more, so they get social surplus. We have seemed to have moved away from the original understanding of how consumer protection worked, which was that it was immoral for firms to extricate all of the social -- we seem to have moved away from that model, and that is fine.

But I think it is the advantage-taking that I really would worry about, and that is the kind of thing I want there to be hard questions asked about.

MS. LOPEZ-GALDOS: Yeah. What I think is that the questions we are addressing here, like from the liability question and the answer from a total perspective to this question right now is that there are no new issues. Discrimination, price discrimination has existed forever. It does not matter whether a machine makes the decision or not, the debate is the debate. We should analyze whether we still -- whether price discrimination, for example, is procompetitive or not or on the other side whether consumers are being harmed or not, which approach we want to take. But it is a debate that we should be having and we have been having even without machines.

So I think we just need to continue talking about these things, but I do not think it makes a difference whether a machine makes a decision or a human being makes that decision.

MR. GILLULA: So actually, I want to disagree. There is something fundamentally different. And if you lump in AI and big data and predictive analytics altogether, then I agree there is nothing new separate on AI. But a major difference is that now there is a -- you are making a decision based on a tremendous amount of data that has been collected as opposed to just like, say, one data point that you happen to notice or one data point you got like the zip code or how many nights you want to stay for the flight or something like that, something that is very clear.

Now, you can potentially make price discrimination decisions based on what websites the person was visiting. Were they looking at budget travel websites versus high-end travel websites? And then there is the question of what happens if -- how were the price discrimination decisions made if you do not have any data on the person? And do they suffer a penalty for preserving their privacy?

If I use a tracker blocker app on my phone and I go to your website and I try to buy a plane ticket and you do not have any history, am I automatically categorized as I have to pay the highest price or not as a punishment for not giving you data about what level I might be willing to pay? So I think that is a difference as opposed to say, you know, what we have been doing for generations. It is not different versus what we have been doing for the last 10 or 15 years.

MS. WORTHMAN: Following up on that, though, is there -- even though these are problems that we have faced before, are there any particular harms that are new based on price discrimination from AI or that is a result of AI? Any new types of harms or is this just the same thing that we have seen before?

MR. CALO: Well, I think there is a huge difference. I think that -- again, I do think you have to group together a bunch of different phenomenon. It is not AI particularly. But, you know, look, for a long, long time, companies have noticed that -- and not just companies like mom and pop shops, everybody has noticed, that there are just cognitive limitations that we all have, right? We just have these limitations to our rationality and that is why everything costs $9.99, right? I mean, obviously, okay?

There is a set of cognitive limitations that behavioral law and economists, Ariely, Kahneman, and so on have -- Christy Jolls at Yale -- have been surfacing over a long period of time. And these are things like prospect theory and status quo bias. And sometimes the FTC actually intervenes and says, you seem to be using status quo bias here with these rebates. We are going to intervene because it does not seem to be fair and you do not seem to understand what is going on.

The issue is that even with all these behavioral economists thinking about how we have cognitive limitations, the list of cognitive limitations is about 45 long, okay?. What artificial intelligence permits you to do because it is so good at pattern matching is to model what rational consumer behavior would look like in a particular environment and then look for deviations that are particular to you, even if they are explicable. Turns out when you are watching “Buffy the Vampire Slayer” on Tuesday night, you are going to pay more for ice cream. I know I am. But the point of the matter is that there will be situations that are very, very specific to you and perhaps not even have a theory behind them.

But what it allows is the mass production of bias. That is what it allows. It allows these systems to figure out where you are specifically susceptible. And, indeed, we see early signs of this already. I mean, you heard earlier a presentation about how Netflix is showing different people posters for shows based on guesses about their demographics or their qualities. You know, that is part of the phenomenon that in the literature is referred to as persuasion profiling, the idea that not just that you be matched to your interests, but that the messages to you to sell you things would be matched to your unique vulnerability.

So, for example, for some reason in your life you are really worried about scarcity, well, that advertisement will say, “while supplies last,” right? And this is the kind of move that marketers are making and it is only possible because of the way that we are mediated by digital technology and we have these intense analytic capabilities and, respectfully, I think that is am enormous distinction from what has come before.

MS. LOPEZ-GALDOS: So obviously, before, we did not have self-driving cars and now, apparently, we are going to have self-driving cars. So we are going to see new things happening. Now, a self-driving car might just cross over a person. What I was trying -- the point that I was trying to make is that the thought process of analyzing the problems and analyzing who is at fault, what was the causality, I mean, the thought process is the same. The same that exists without human beings is just applied to the new setting.

I think the theories and thought process should remain -- we should not think in the abstract. We should think like we have a lot of analysis in tort law, for example, and we want to say who is responsible, who is not. In a self-driving car, there is software, hardware, there are apps, there might be somebody inside the car that was doing something as well. And what I mean is that in the thought -- when we are analyzing who is at fault and who is liable for crossing over two people, the thought process of, for example, causality should be the same as without AI. That is an example -- for example, of the point that I was trying to make.

MS. GEORGE: So I am going to ask one final question and then I think we are going to wrap up. It is going to be a compound question. Because I like that.

So are there ways in which the FTC should expand or rethink the notions of unfairness and deception when it comes to AI and what educational role should the FTC play with these new technologies, both for consumers and businesses?

Marianela, do you want to start?

MS. LOPEZ-GALDOS: I think it is a very good final question. I think the FTC is doing a great job in putting together these hearings, as I said in the beginning. I think AI is just a machine learning -- it is at a nascent moment. I think it is very important to keep having a dialogue with businesses, with the community, with the consumers, with experts, and see where we are going to and see whether there is anything that needs to be refined, for example, of existing laws or not.

But what is very, very important is not to think in the abstract of AI. We talk about AI as if -- you know, at this moment, there are marvelous things that can be done. I think there is a lot of potential, but I really think that before stepping and regulating or saying, oh, this is going to be a disaster, everything is going to be mass-biased, et cetera, we really need to understand where we stand, what engineers can do, what companies are working on.

I think companies, at least the ones that CCIA represents, are willing to cooperate with the authorities, are willing to engage in adopting principles. And I think having an open and frank dialogue about what is going on is key to make sure we get the right approach. So society can really profit from AI.

MS. GEORGE: Irene, you just want to continue down the line?

MS. LIU: Sure. Again from the beginning, I feel that the FTC framework and the existing laws are sufficient and the fact that it is broad enough that it can capture AI, I think that is great. I think FTC has withstood the test of time because it is broad.

But at the same time, I do think -- I agree with Marianela that it is important for the FTC to continue talking to the industry, also with other regulators and academics to make sure that they are abreast of this nascent technology.

There is also movements across the globe, it is not just a U.S. phenomenon, but just globally. Again, there is a recent universal guideline for AI that was launched in 2018 by a number of data protection officers recently. The World Economic Forum is working on this issue. Regulators in Europe, China, have taken a deep interest in AI and so there is a lot of cross-country developments within AI as well that the FTC can also engage in to make sure that it stays ahead in terms of the policy developments around the world so that we are not hindering innovation, but fostering it as well. So from that perspective, I think the FTC Act is moving in the right direction with these types of hearings as well.

From an education standpoint, the FTC can also play a role in educating consumers to understand what is AI. Again, because it is a new technology, people hear about it. We talk about it all the time in Silicon Valley, but it may not be known to the rest of the country. So just educating people about what chatbots are, what it means when you are choosing Netflix on a Tuesday night and watching “Buffy the Vampire Slayer,” what the impact might be. It might be that your ice cream prices might go up or it may be that your Netflix fee might go up if you are a more avid watcher than others.

So just understanding the impact of the data would be helpful to consumers and also encouraging companies to implement AI not just to exploit data, but to think about it holistically is really important and encouraging companies to do that from that framework of advancing society versus exploiting the data is something that FTC can take on, too.

MR. GILLULA: So I am actually going to answer the question in reverse order. In terms of consumer education, I think to accomplish that mission, the FTC needs a much more robust staff of technologists. They have only somewhat recently started having technologists on staff. I feel like the FTC should have as many technologists as lawyers at this point. And, obviously, that is not where we are.

I also realize that is not in the FTC’s ability to change. So if you are a Congressman or a

Congresswoman sitting in the audience, this is my plea to you is increase the funding for technologists at the FTC because those technologists can help with explaining AI and what to expect in a consumer standpoint to consumers. They can also help explain it to the lawyers at the FTC when they are doing enforcement actions or they are doing investigations. They can help explain it to policymakers. So I think there is a real need for a really robust technical staff there.

In terms of whether or not the FTC Act sufficiently captures everything that we might worry about with regards to AI, I still worry a little bit about the fact that -- I mean, I guess there are two parts. One is whether -- I mean, at least -- and, again, you got the only nonlawyer I think on the panel talking. The FTC Act -- when you are talking about harms and unfair and deceptive, you are talking about what is the cost-benefit analysis. And I worry a little bit that when we are talking about privacy, in particular -- so, again, this comes back to rolling AI and big data and predictive analytics into the same thing.

But when you are talking about privacy, what may be good for society is not necessarily good for the most vulnerable part of the population because privacy is really about privilege. You know, a cis -- I am a cis, white guy, middle class, like I am boring. Like you could know everything about me and it does not matter because I am not worried about something happening to me. But for many people with very different demographics, they are very worried about what data gets out about them.

And so while, on average, when we are making that sort of cost-benefit analysis about what works for society, that might make sense. But when we are talking about privacy, we really need to be thinking about what works for the most vulnerable part of the population.

MR. CATE: So I think the FTC has enormous capacity under Section 5 and FCRA and so forth. And as Ryan was saying earlier, I think it should be asking the hard questions and flexing those muscles. Having said that, I actually do think additional legal authority is likely to be necessary. Some of that may be based more on what we might call procedure, but in terms of ways that companies go about making decisions and documenting those decisions about the use of database automated decision making that affects individuals in significant ways.

And then I do not think there is actually a shortage of information, I think we have too much information right now about AI and that one role that the Commission might very productively play, as it is doing now, is helping to sort of sort through some of that information. I mean, everyone on earth now has a code on AI. They all start with fairness and have no idea what fairness means, not the first idea.

And so helping to -- for example, as you have begun today, thinking through what is fairness, what are the elements of fairness, how do you measure it, what is a desirable goal. The same thing about harms. I do not think we have any agreement at all about what are harms. I mean, we know the extreme of harms. If someone is specifically injured or they lose money, that is a harm. But what about between where we are and there?

So in this area, I think the FTC has an enormously important role to play and, frankly, a great deal of experience to draw on in trying to kind of sort through all of the stuff that is out there and emerging and try to help make sense of it for individuals and for businesses alike.

MR. CALO: Yeah, I mean, there has been a lot of healthy back and forth and disagreement about certain things on this panel, but I think that you are seeing a rough consensus that the Federal Trade Commission is well suited both because of its expertise and because of its century of protecting American consumers. I think we need an FTC that is very assertive and uses the full range of its powers and pushes the definition of unfairness and deception and updates it for contemporary context. That is what is so beautiful about a standard is that it can be updated. And if these new technologies are as powerful as people claim, so powerful that we need to get out of their way, then they are also the kind of thing that require a change to law and legal institutions.

So my hope, too, is for -- I do not know that there is any additional authority really needed. I just think that the Federal Trade Commission should be emboldened to pursue these very assertively and that Congress and the courts should let them do their job.

MS. GEORGE: With that, I want to thank our panelists and audience for an exciting discussion.

I want to remind everyone to come back for day two tomorrow for more interesting insights. And thank you all for participating in this process. Thank you.

(Applause.)

(Hearing adjourned.)

CERTIFICATE OF REPORTER

I, Linda Metcalf, do hereby certify that the foregoing proceedings were digitally recorded by me and reduced to typewriting under my supervision; that I am neither counsel for, related to, nor employed by any of the parties to the action in which these proceedings were transcribed; that I am not a relative or employee of any attorney or counsel employed by the parties hereto, not financially or otherwise interested in the outcome in the action.

LINDA METCALF, CER

Court Reporter