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

1 FEDERAL TRADE COMMISSION
2 I N D E X
3 PAGE:

22
23
24
25

1 WELCOME AND INTRODUCTORY REMARKS
2 MR. GAVIL: Good morning, everyone. My name
3 is Andy Gavil, and I'm a Professor here at the Howard
4 University School of Law. On behalf of Dean Danielle
5 Holley-Walker, my faculty colleagues, and our
6 students, I'd like to welcome the FTC and all of you
7 to Howard for Hearing Number 7 of the FTC's hearings
8 on Competition and Consumer Protection in the 21st
9 Century. We are very happy to cosponsor today's
10 event, and I want to thank the FTC and the many people
11 at the agency and here at Howard who have worked hard
12 over the past few months to organize these hearings.
13 As you all know, today's topic is
14 Algorithms, Artificial Intelligence and Predictive
15 Analytics. As is immediately evident from both the
16 list of questions the FTC has posed and the agenda for
17 today and tomorrow's programs, these hearings have
18 been purposefully designed to take a broader and more
19 interdisciplinary perspective than any of the previous
20 ones.
21 Moving well beyond the usual collection of
22 academic and practicing economists and lawyers, FTC
23 staff have assembled an impressive collection of
24 academics, public servants, technologists, scientists
25 engineers, and industry leaders, but, of course,

1 there's still lots of lawyers and economists.
2 The goal is to educate the agencies and the
3 broader competition and consumer protection policy
4 community so we can all obtain a better understanding
5 of the technologies that are transforming our economy,
6 as well as our political and social environs. We'll
7 hopefully learn more so we can better understand the
8 business models and practices of our time and so we
9 can differentiate myth from reality, promise from near
10 and long-term prospect.
11 The ability to take on this kind of
12 prospective study is a hallmark of the FTC and one of
13 its great institutional strengths. It is especially
14 fitting that such a forward-looking approach is being
15 taken here at Howard. Only two years after Howard
16 University was chartered by Congress in 1867, this law
17 school was founded with the aspiration of producing
18 lawyers who would lead the future fight to realize the
19 full promise of the reconstruction amendments to the
20 Constitution of the United States.
21 Next year, we will celebrate our
22 sesquicentennial, and for that occasion, instead of
23 looking backward, we have selected a theme that looks
24 forward, “The Next 150." As is true for the FTC and
25 for today's hearings, any institution that fails to

1 look forward is bound to fall backward.
2 In closing, please note that the event is
3 being photographed and webcast and will be posted on
4 the FTC's website, and that by participating all
5 attendees consent to those conditions.
6 Please also note that our students will be
7 coming and going throughout the day and are available
8 to answer your questions. Please get to know them
9 while you are here and feel free to seek them out if
10 you have any questions or concerns.
11 Finally, it's my great pleasure to introduce
12 our first presenter. Our scheduled presenter, Michael
13 Kearns, has been slightly delayed, so we're going to
14 start with John Dickerson from the University of
15 Maryland, and hopefully Michael will arrive in time to
16 follow John. Again, welcome, thank you, and enjoy the
17 hearings.
18
19
20
21
22
23
24
25

1 PRESENTATION: INTRODUCTION TO ALGORITHMS,
2 ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYTICS
3 MS. GOLDMAN: Hi, I'm Karen Goldman. I'm an
4 attorney in the Office of Policy Planning at the
5 Federal Trade Commission, and I just want to introduce
6 you to John Dickerson, who is an Assistant Professor
7 in the Department of Computer Science at the
8 University of Maryland, College Park. Welcome.
9 MR. DICKERSON: Thank you, Karen. It's a
10 pleasure to be here. I am John Dickerson, I'm a, I
11 guess, third-year Assistant Professor at the
12 University of Maryland and right up the street in
13 College Park, and today I'll be talking about an
14 introduction briefly introducing the audience to
15 algorithms, AI, and predictive analytics.
16 And so for this talk, I'd like to start with
17 a motivational quote which sounds like it was written
18 a long time ago, and that's because it was. So
19 "although machines can perform certain things as well
20 or perhaps better than any of us can, they infallibly
21 fall short in others," by which means we may deduce
22 that they did not act from knowledge but only from the
23 disposition of their organs.
24 And this sounds old because it was written
25 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 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 built on now hundreds of years of philosophy and mathematics, but the

1 general pitch here is that if intelligence can be
2 simulated by mathematical reasoning, that is reasoning
3 is just reckoning, and mathematical reasoning can be
4 simulated by a machine, then can a machine simulate
5 intelligence?
6 So AI, artificial intelligence, the word was
7 coined by John McCarthy in either 1955 or 1956,
8 depending on how you count, it’s '55 in a proposal, to
9 fund the Dartmouth Summer Research Project on
10 Artificial Intelligence. And you'll hear this called
11 the Dartmouth Conference. This occurred in the summer
12 of 1956.
13 And there are some fun quotes in there
14 saying basically we can solve artificial intelligence
15 in three months or we can solve artificial
16 intelligence in one generation, but the one I'd like
17 to pull out is that every aspect of learning or any
18 other feature of intelligence can be so precisely
19 described that a machine can be made to simulate it.
20 So even in the 1950s, 1960s, folks were making
21 statements like this.
22 So a quick spoiler, this hasn't happened
23 yet. We can just shut this down right now. But,
24 progress has been made. So how does that progress
25 occur? Well, this is a cycle of basically R&D

1 progress that you'll see repeating in the AI world,
2 and this has happened since basically 1956, where some
3 new advance, maybe a new technique, new hardware
4 happens. Fast progress is then made on old, hard
5 problems. So it could be a new mathematical
6 technique, it could be new hardware, GPUs, these
7 graphics processing units, are one of the main drivers
8 in the current sort of fast progress being made on
9 problems that we're seeing now.
10 But eventually you start to hit road blocks.
11 And at this point, the academic community, the
12 industrial community starts to get pessimistic, this
13 bleeds into the press, and at that point, everyone is
14 pessimistic about progress, funding dries up, progress
15 dries up and so on. We wait until the next large
16 advance.
17 And so this is the cycle that occurs in most
18 sorts of verticals. It occurs in AI research as well.
19 In AI, though, we call it a cycle of basically AI
20 summers and AI winters. The winters are when funding
21 dries up and nothing happens; the summers are
22 basically what we're going through right now, where
23 we're seeing large advances driven by sort of recent
24 hardware and mathematical advances.
25 So this is a bit pessimistic, this cycle,

1 but like I said, progress has been made. So this has
2 been cycling for arguably maybe six or seven times
3 since the 1950s, but every time we go through this
4 loop, progress is made, new problems are solved, and
5 new problems are encountered.
6 So what is AI? AI, many definitions, the
7 one I'll use here is the ability to process and act
8 based on information via automation. So we can break
9 this down roughly into four segments. One is
10 perception. I want to be able perceive the world
11 around me. That could be the physical world; that
12 could be the virtual world. I want to be able to
13 learn something about it. So I get some signals about
14 the world, then I learn something about them. Maybe I
15 learn a model.
16 I want to abstract and generalize that model
17 so that I can use it in other situations. And what do
18 I mean by use? Well, maybe I can reason about this
19 information, I can reason using my model and then act
20 within the world. Again, that could be virtual, that
21 could be physical.
22 So if I can create this automated system,
23 roughly, I have created what we would call AI. So
24 let's keep moving through this history of AI until we
25 are where we are today. Roughly we can split AI

1 research into some first-wave AI, second-wave AI, and
2 then maybe 2.5 or third-wave, which is where we are
3 right now.
4 In this first wave, primarily, researchers
5 focused on what is called search. So this is either
6 searching through a potential solution space, some
7 quick examples, chess is a good example here where we
8 had, say, Deep Blue beating Kasparov via a
9 sophisticated algorithm that did search through using
10 domain-specific heuristics, expert knowledge, for
11 instance. Folks who played a lot of chess encoded
12 heuristics into the search algorithm; it would search
13 through the solution space to find, say, the next move
14 to play.
15 Now, another hallmark of first-wave AI is
16 something called expert systems. And this also relies
17 on basically bringing in a lot of expert knowledge to
18 form some sort of large database of rules, of
19 knowledge, of facts about the world, using some sort
20 of inference engine, typically based on logical
21 reasoning, to make new sort of conclusions based on
22 these facts, and then some sort of action, I/O
23 system to interact with the human. So this is
24 basically the world up until maybe the ‘80s in
25 terms of AI.

1 Now, there were some large successes here,
2 so one example that I used earlier is this chess
3 champion falling to basically a sophisticated search
4 algorithm. And there are many more. And, in fact,
5 techniques from first-wave AI are still used in
6 practice, but they're decidedly brittle and they
7 really don't have any real learning capability. So
8 they're really sort of a function of just the
9 knowledge that you encode into them.
10 There's a huge overhead to encoding that
11 knowledge. Right, I have to ask, say, every member of
12 the audience and everyone watching to tell me all the
13 facts that they know about the world and then I have
14 to store that somehow, and that might be brittle and
15 that might not generalizable. They're very, very
16 brittle systems, but they do allow me to do in-depth
17 specific reasoning. Right, if I ask a bunch of
18 experts for facts on a specific vertical, then I can
19 do a lot of fast automated reasoning about just that
20 vertical. So that can be good, but it's very
21 difficult to generalize.
22 And if you recall back to that earlier
23 slide, we want generalizability, we want abstraction
24 because we want to create some system that's able to
25 encounter new environments and still act in a

1 reasonable way.
2 So in terms of those four boxes, first-wave
3 AI generally does perception reasonably well in the
4 sense that I have asked all audience members to give
5 me facts, and it can do reasoning and acting
6 reasonably well, but it won't learn and it won’t
7 generalize.
8 Now, there were some transition points in
9 multiple areas of sort of AI research. One of these
10 is something called natural language processing, which
11 says, can I get a computer to ingest, say, raw text or
12 can I get Alexa to ingest signal from your voice and
13 then have it understand that in some sense. So in
14 natural language processing, up until about the late
15 1980s, most of the rules for doing this sort of
16 translation or understanding were done via hardcoded
17 sort of expert rules.
18 Around the late ‘80s, probabilistic models
19 started to come into play. Okay, so this is going to
20 sound more like machine learning like folks have maybe
21 heard about in the press. These are models that
22 ingest, in this case, large text corpora and learn
23 patterns in that data.
24 To look at a different vertical in AI, so
25 autonomous vehicles rely heavily on something called

1 computational vision, which says, hey, I have a video
2 image, can I understand what's going on in that image.
3 Say I'm a car and I'm driving along, and I have a
4 still image of the road in front of me, can I
5 understand that there's a stop sign and a pedestrian
6 and dog in front of me and so on. So in autonomous
7 vehicles, in the mid 2000s, ARPA ran what they call
8 a Grand Challenge, in fact their first Grand
9 Challenge, which asks, can I create a vehicle that
10 can drive some hundred-plus miles across the desert
11 autonomously?
12 In 2004, no vehicles completed this task.
13 In fact, I think the longest trip that a vehicle took
14 was something like ten miles. And these vehicles
15 relied heavily on hand-coded rules that say something
16 like, in general, when you're, you know, ten degrees
17 away from the sun and you're driving forward at a
18 particular speed, then a shadow is going to be a
19 shadow instead of a rock with some set of features
20 associated with it. And, again, this is a very
21 brittle system. This is not going to generalize very
22 well.
23 But then in 2005, five teams completed the
24 entire trip, so 100-plus miles. And this is because
25 they started using these probabilistic models. And,

1 in fact, you can see the general manager for the
2 program, Strat at the time, had a fun quote:
3 "[Vehicles] were scared of their own shadow,
4 hallucinating obstacles when they weren't there."
5 And this is for those prior systems. And then
6 probabilistic models allowed them to get around this.
7 So you can see similar transition points
8 throughout all core AI areas, in the late '80s, in the
9 '90s, up and through basically the mid-2000s. And
10 this happened because of three things. One is
11 computational power increased, and this is the story
12 of basically computation since the '40s or '50s. This
13 has played a driving role in AI development as well.
14 Number two, storage costs decreased. I
15 don't have to pay a lot of money to store a lot of
16 data. And, three, everyone in this world now relies
17 on statistical models, maybe with some expert input,
18 but still statistical models.
19 So this takes us into the second wave of AI,
20 and there's no hard date for this because it happened
21 differently in different verticals in this world.
22 Here, we're relying on this assumption now that we've
23 learned the hard way, multiple times, that encoding
24 all knowledge explicitly does not work. It doesn't
25 scale. It's very brittle and it's very difficult to

1 handle uncertainty.
2 The new idea is that we should create a
3 general statistical model for a problem domain. We
4 should create a statistical model for natural language
5 or for a type of natural language or for autonomous
6 driving, a type of autonomous driving. What do we do
7 with that model? Well, we feed in data from the real
8 world or maybe simulated data until it looks right.
9 And this is going to be characterized by statistical
10 learning.
11 So the reason why these models have taken
12 off is because if we input a different data set or,
13 say, set of data sets into these models, we'll learn a
14 different model and then we can deploy that in a
15 different environment. So it's much more
16 generalizable.
17 Now, some examples. In machine translation,
18 for instance, going back to this natural language that
19 we discussed earlier, we can feed in multilingual text
20 corpora to learn relationships between language. So
21 say we want to translate French to English, one of the
22 early multilingual text corpora came from Canada,
23 where there are rules stating that, say, any
24 government ruling has to appear both in English and
25 French. And so now we have a mapping between English

1 and French documents, we can feed that into a model
2 and we can learn a way to translate between the two
3 systems.
4 Autonomous vehicles. We can feed in videos
5 and tests of successful driving into a model and then
6 learn what scenarios are safe or not safe or maybe put
7 some error bars around what scenarios are safe in
8 general.
9 Face detection, face recognition. I can
10 feed in many labeled faces of people. Here is where
11 the face is, or here is where the face is and an idea
12 associated with that, to learn what a face looks like
13 or to learn what, say, your face looks like.
14 So these types of models are very good at
15 perception, and they're very good at learning.
16 Remember, we're training these models, these general
17 models, based on a data set, and if we feed in a
18 different data set, we're going to get a different
19 result, so they’re reasonably good at abstraction and
20 generalization as well, so long as your model is
21 general enough and so long as you have enough data.
22 But there is no reasoning or acting. I've made no
23 statements about, say, when one should turn the car in
24 -- turn the wheel in the autonomous vehicle.
25 So a quick example model. Remember, these

1 are systems that rely on statistical learning to train
2 probabilistic models that will tell us something about
3 the world. A quick example is a neural network. So
4 these appear a lot in the news now, which is why I've
5 chosen them, but they're not a new idea. Indeed, that
6 1955 proposal where McCarthy defined AI, used the term
7 AI for the first time, also discusses neural networks.
8 I believe they were called neuron networks at the
9 time. So this is not a new idea.
10 The general idea of neural networks is that
11 one should pass information into this input layer,
12 which you see on the left side of the screen. So that
13 information could be pixels of an image. That
14 information could be something with audio signal. It
15 will cascade through the network, along basically a
16 series of pipes that go through nodes, and these pipes
17 have, say, different widths that can be controlled by
18 a learning algorithm.
19 And then the final layer of this network
20 that has information flowing through it will create
21 some sort of guess. In the case of, say, classifying
22 images, here we have cats and dogs, it's going to
23 create, say, a probabilistic model of whether or not
24 an image is a cat or a dog. And that gives you some
25 signal as to how good or bad your statistical model --

1 in this case a neural network -- is acting.
2 A very general model, so long as we can feed
3 information into it via that input layer and so long
4 as we can judge the output and so long as we can
5 actually learn, so make the network better, using
6 sophisticated optimization techniques, we can use this
7 for many problems and, indeed, that is what we've
8 seen, so long as we can, again, train these models
9 through repetitive sort of optimization algorithms.
10 So another sort of buzzword that one sees in
11 the press a lot is a deep neural network. Again, not
12 a new idea. These existed, I think, since the 1980s,
13 and they're just these neural networks that we had on
14 the last slide but with more, quote, unquote, hidden
15 layers. These are the layers in between that input
16 and that output. So I can add more and more of these.
17 I can create more piping -- intricate piping between
18 these different nodes to learn new patterns in the
19 data.
20 And sometimes, indeed, we can stack many,
21 many, many, many, many more nodes, so order of
22 hundreds of thousands, millions, et cetera. So these
23 are very large models. And, again, this is because we
24 have increased computational power and cheap storage.
25 That idea for deep networks has existed

1 since the ‘80s, but we've seen them taking off in the
2 last five to ten years because of advances in
3 hardware, because of a huge increase in the amount of
4 data that exists. So we have large firms collecting
5 data; we have the government collecting data; and we
6 can now store it cheaply, access it quickly, and
7 because, indeed, from the R&D community, there have
8 been much better methods developed for learning
9 basically how to make a good one of these.
10 They're hugely successful. They're good at
11 detecting anomalies in data, for instance, credit card
12 fraud. They're good at voice recognition. You've
13 seen Alexa, Siri, Google Assistant, et cetera.
14 They're great at machine translation, language
15 generation, game playing. Some recent high-profile
16 success stories such as AlphaGo playing basically
17 expert-level, Go, DeepStack Plane, expert-level Heads-
18 Up Poker.
19 Self-driving cars are starting to take off.
20 Video search, audio search, finance, et cetera. These
21 are all success stories in part due to deep learning.
22 Not a new idea, driven by advances in hardware and
23 training them.
24 Nobody understands why they work very well,
25 and this is a common story in AI as well and this is

1 something that we're seeing more and more appearing,
2 which is humans have sat down, they've designed the
3 network structure, they've designed what those nodes
4 and what those connections between the nodes look
5 like. Maybe they're encoding some domain expertise.
6 There are some known heuristics that you can rely on.
7 There's a trial-and-error process, and maybe actually
8 other AI is actually coming in and trying to train
9 these models or structure these models in a better
10 way, but nobody knows when or why they don't work in
11 general.
12 So they work well in expectation, which is
13 why we see machine translation systems, which is why
14 we see Alexa and Siri in households now, but when they
15 fail, it can be very confusing, it can be reasonably
16 catastrophic, and it can be very hard to explain.
17 And some recent research pushes funded by
18 the DOD, funded by industry, funded by nonprofits,
19 have started noticing that, hey, an adversary can
20 exploit this kind of behavior. When I have a system I
21 trust most of the time but it can be exploited in very
22 odd ways and I don't understand why or when that
23 happens, then I can wreak some havoc in these systems.
24 So I'd like to take a step back. So now
25 we've talked about deep learning, we've talked about

1 machine learning, and we've talked about AI. And,
2 roughly, AI is this sort of four-pillar approach to
3 perceiving the world, learning about it, building an
4 abstract and general model, and then using that to
5 act and reason. Machine learning is just one way to
6 build these models, where we do not focus on acting
7 and reasoning but we focus on perception, on learning,
8 on abstraction, and on generalization. And deep
9 learning is just a specific form of basically
10 representational learning, so it's a type of machine
11 learning.
12 Right, so every time you hear deep learning
13 in the news, you can replace it with machine learning
14 mentally. It's just a way to solve a machine learning
15 problem.
16 So some present-day movements in AI,
17 understanding bias and methods for debiasing. You'll
18 hear about this I think throughout today and tomorrow,
19 many of the topics on this slide. So this is sort of
20 a teaser. Understanding bias and methods for
21 debiasing. So if I feed skewed training data into
22 these systems -- remember, these are statistical
23 models that are trained on data from somewhere in the
24 world. If I feed skewed data into the system, then
25 I'm going to learn something that represents that

1 skewed data. So how do we understand when that
2 happens and can we create systems that still feed in
3 this biased data which might be the only data that
4 exists but spits out a model that is debiased?
5 As mentioned before, adversarial reasoning
6 in multi-agent systems, learning to act with
7 cooperative actors, learning to act with adversarial
8 actors, so bringing in older fields such as game
9 theory into these new methods for solving those
10 problems. How do I say design -- well, I'll talk
11 about this in a few slides, but how do I design
12 policies as a firm to compete with other, say, firms
13 that are both cooperative and adversarial? Can I do
14 this based on machine learning?
15 Also mentioned on the previous slide,
16 robustness to noise, robustness to adversarial
17 attacks, both in terms of theoretical robustness and
18 empirical robustness. How do I design automated
19 systems that fail less, that are robust to attacks and
20 that fail more predictably, because obviously these
21 systems will always fail at some point.
22 And in that vein, explainable AI, there's a
23 lot of money going into this as well because it's very
24 difficult to interpret the results that come out of
25 these systems from time to time, so can we produce

1 human-understandable models that also work well?
2 And one final move in the AI community has
3 been reinforcement learning. It's a type of machine
4 learning, but it's a type of machine learning that
5 also focuses on learning to act and reason. So now
6 we're getting closer to that initial definition of
7 artificial intelligence. Here we have an agent, maybe
8 physical, maybe virtual, that's going to act within an
9 environment. It's going to receive a reward signal
10 and then maximize total reward. It wants to find the
11 actions to take for any state in the world such that
12 when it takes that action, it is treated well in the
13 future, it receives reward and expectation in the
14 future. And I'll give you some examples of this at
15 the end of the talk.
16 So here again, again, reinforcement
17 learning, not a new idea, but deep networks have been
18 used extensively here to revolutionize their use and
19 practice. So here we have deep networks that are used
20 to, say, reduce the complexity of representing the
21 environment. Remember, I can't actually write
22 everything down, I don't want to represent every
23 single fact in my computer, so now I'm going to learn
24 some abstraction of the world and then act on that.
25 So reinforcement learning is taking us

1 closer to what we want to call AI. We have
2 perception, we have learning. These are just like
3 machine learning. We have abstraction and
4 generalization, again, moving toward that. Again, if
5 we train these models on different data, we get a
6 different trained model, and we're starting to move
7 toward reasoning and acting here.
8 So in the context of this audience, I
9 thought I would do maybe a quick deep dive into a few
10 uses of AI, particularly in something called market
11 design. So markets provide agents the opportunity to
12 gain from trade. Many markets require structure to
13 operate efficiently. Market design is going to tackle
14 this problem via what's called economic engineering.
15 So I put on my economist hat and I put on my
16 engineer’s hat and I put on my mathematician's hat.
17 I’m wearing three hats at this point, but I can use
18 these hats to design a market, how do I structure the
19 market, how do I constrain the market such that I
20 achieve some sort of efficiency goals.
21 AI is increasingly helping with the design
22 of these markets. For instance, these automated
23 methods can use data to help designers characterize
24 families of market structures. They can be used
25 obviously for predictive methods that anticipate, say,

1 future supply and demand in electricity markets or
2 finance markets.
3 One example, as a computer scientist, this
4 is close to my heart because a lot of the money in our
5 world comes from this, is using AI in online
6 advertising. So online advertising markets generally
7 match advertisers with consumers. Many billions of
8 dollars, and this is an increasing market, many, many
9 billions of dollars are being used here, and it's a
10 driving force in the technology sector.
11 Machine learning models in this space right
12 now are being used to divide customers into very fine-
13 grained and automatically generated segments. So no
14 longer just male/female but something far, far more
15 fine-grained than that. That’s learned automatically.
16 They're being used to set reserve prices and auctions
17 based on user modeling and bidder behavior, again
18 automatically.
19 They're being used to automatically generate
20 the creatives, that is, the artwork that you see pop
21 up on your screen, to automatically generate those,
22 say without human input, to fit a specific customer’s
23 predicted wants. All automated.
24 Reinforcement-learning-based tools --
25 remember, this is that form of machine learning that

1 also focuses a bit on acting -- are being used to help
2 advertisers, for instance, bid automatically on these
3 very fine-grained segments. Remember, now we have,
4 say, millions of segments. How do I bid on that, I
5 can use a machine-learning-based model to do this.
6 Another example, AI in electricity markets.
7 Here, matching supply and demand is extremely
8 important. It relies heavily on demand forecasting.
9 Machine-learning-based techniques are going to provide
10 very accurate demand forecasting, which leads to very
11 stable market prices and more efficient power usage.
12 Reinforcement-learning-based techniques --
13 remember ML plus some form of acting -- are going to
14 allow us to activate or deactivate expensive
15 heterogeneous power sources to maintain that
16 stability. So I can predict better demand, I can
17 predict better demand at particular time points
18 further into the future, and then I can make a plan to
19 boot up or boot down particular power sources such
20 that I maintain market stability, such that I reduce
21 brownouts and so on. Again, automated.
22 And my final example is AI and kidney
23 allocation. This is close to my heart. I've done a
24 lot of work in this space. So here, kidney exchanges,
25 for instance, are an organized market where patients

1 with end-stage renal disease enter and are able to
2 swap donors -- willing living donors -- to receive new
3 organs.
4 It's a really interesting paradigm that's
5 been around for, say, 15 years now, and it accounts
6 for something like 10, 11, 12, 13 percent now of all
7 U.S. living donations of kidneys. Hundreds of
8 transplant centers are involved in this organized
9 market, in fact, multiple organized markets.
10 And, here, AI-based tools are also
11 operating. Now, this isn't fully automated, but
12 they are, for instance, semiautomatically and
13 optimally subject to human value judgments, matching
14 donors to patients, both in the U.S. and also
15 worldwide. Here, I've called out the United Kingdom
16 and the Netherlands, but in many countries. They're
17 providing sensitivity analysis at a level that humans
18 cannot for new policies. And they're learning from
19 data the quality of, say, potential matches in this
20 market.
21 Now, let's return to some open questions
22 and some recent pushes which will, I guess, trigger
23 good discussion for the rest of today and tomorrow.
24 So one is, how and why does deep learning work? So
25 I've mentioned not a new idea. Neural networks

1 existed since the ‘50s; deep learning existed since
2 the ‘80s. Now we have new hardware and now we have
3 new training techniques, these tend to work very well
4 in expectation, but when they fail, they fail
5 confusingly. Why do they work?
6 How can we handle incentives of competing
7 agents? All those three market examples that I showed
8 you, firms are obviously going to compete against each
9 other in this space. The government, regulatory
10 agencies have their own incentives as well.
11 Individual participants have their own incentives.
12 How do we handle this, how do we encode other aspects
13 such as fairness, accountability, and transparency
14 into these systems?
15 How do I ensure that my automated system
16 doesn't marginalize, say, an already marginalized
17 class in the ever sort of increasing hunt for
18 efficiency? How do I even define this? How do I
19 define fairness? This is a classic question in
20 economics that computer scientists are now starting to
21 struggle with as well. How do I implement this in a
22 scalable way, in an understandable way?
23 Ethical AI, this will be talked about, I
24 believe, later, by folks like Henry Kautz, how do I
25 divide the labor between policymakers, such as those

1 in this audience, who are ethically trained and
2 ethically minded and technically trained, perhaps
3 ethically neutral AI and machine learning experts?
4 So I can implement, say, a very sophisticated system,
5 but I need to then produce some sort of aggregate
6 output that I can pass back to policymakers to ensure
7 that this is reflecting the aggregate human value
8 judgments of those who control the systems. How do I
9 do that? And there are close ties in this sort of
10 exploration to the world of privacy and the world of
11 social norms.
12 So in general, our end goal is to create
13 these systems that perceive the world, learn from it,
14 create some sort of generalizable model and then
15 inevitably learn to act using that model. We're not
16 quite there yet, but there's a lot of hope in this
17 space. But I'm going to say that maybe this isn't
18 even the biggest problem. The biggest problem is
19 going to be the interplay between these systems and
20 society, ethical issues, societal norms, human value
21 judgments. How do we play between, say, these sort of
22 sophisticated machine-learning-based approaches to
23 what I've shown here on this slide and the rest of the
24 real world? So I'll leave it at that.
25 MS. GOLDMAN: Thank you very much, Professor

1 Dickerson, for that excellent introduction to the
2 field and for the questions that will be coming
3 throughout this hearing.
4 (Applause.)
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 OPENING ADDRESS
2 MS. GOLDMAN: As I mentioned, I'm Karen
3 Goldman. I'm with the Office of Policy Planning at
4 the FTC, and now I'd like to introduce our next
5 speaker, Michael Kearns, who is a Professor at the
6 Department of Computer and Information Science at the
7 University of Pennsylvania. Welcome, and we're
8 looking forward to your address.
9 DR. KEARNS: Okay, thank you. So not
10 only am I late, but I also missed the deadline to
11 give slides last week. So there's two strikes
12 against me already. But hopefully the time I would
13 have spent hacking PowerPoint I put into thinking
14 instead, and so I'm just going to speak informally
15 from notes.
16 So as she said, I'm on the computer science
17 faculty at the University of Pennsylvania. My main
18 research area is machine learning and related areas.
19 I've been in this area since I was a doctoral student
20 in the 1980s, before machine learning was a thing in
21 society. And so I was just asked to give some
22 introductory framing remarks based on the agenda
23 that I saw, which contains, like, lots of topics
24 that are near and dear to my technical and related
25 interests.

1 In particular, following on the last
2 speaker, in the last few years, I've been thinking
3 quite a bit about ethical and social issues in the use
4 of machine learning and algorithmic decision-making
5 more generally. And I also saw that there are some
6 discussion or a panel about sort of competition and
7 marketplace questions introduced by machine learning.
8 I hope to make some less technical remarks about that
9 because I think that's less scientific to say there
10 but a lot of interesting things to discuss, and also
11 relatedly topics related to consumer protection and
12 abuses by machine learning and AI.
13 And so what I just want to do with my time
14 is make some informal remarks, provide some personal
15 opinions on these topics based on my own experiences
16 and research, and, you know, hopefully cue things up
17 for the next couple of days for the rest of the
18 speakers.
19 So as the last speaker mentioned, there has
20 been a lot of discussion not really first in the
21 technical community but first in the mainstream media
22 and society at large, about the many things that can
23 go wrong when applying machine learning and AI and
24 related methods to algorithmic decision-making.
25 And before I describe -- say a little bit

1 more about what can go wrong and what we might hope to
2 do about it scientifically, I thought I would start by
3 sort of just framing how things can go wrong in the
4 first place. And so one thing you might wonder is,
5 you know, if there is a lending model or a consumer
6 credit-scoring model that exhibits racial bias, for
7 instance, or there’s some data analysis or machine
8 learning methodology that leaks personal, private
9 consumer data, you might -- it's a reasonable thing to
10 wonder whether this happens through active
11 malfeasance. You know, are there evil programmers at
12 tech companies who, you know, put in a line in their
13 code that says if the person’s race is this then do
14 this; if it's some other race, then do something else;
15 or whether they program back doors into their code
16 that permit privacy leaks.
17 And there’s good news and bad news here. My
18 strong belief, and I think those people who work in
19 the field would say that, no, there is absolutely no
20 such malfeasance going on by evil programmers at
21 technology and other companies. So that's the good
22 news. The bad news is that the truth might actually
23 be a little worse, which is these sort of collateral
24 damage or consequences are actually the natural
25 byproduct of applying the formal, fundamental,

1 scientific principles of machine learning and AI. And
2 I'll say a little bit more about that.
3 The vast majority of what I think we would
4 call algorithmic decision-making is actually a little
5 bit more specifically almost always driven by machine
6 learning these days. So, in particular, when you
7 think about the algorithms that make things like
8 lending decisions or decide what ads to show you on
9 Facebook or Google, these generally are not what you
10 should think of as hand-coded or programmed
11 algorithms, but rather they're the result of taking
12 data, you know, historical data, whatever that means
13 in a given domain, giving it to an algorithm, and that
14 algorithm, of course, trains a model on the data. And
15 then at the end of the day, it's the model that's
16 actually making the decisions. It's the model that's
17 actually deployed in the field.
18 And, typically, the algorithm that
19 transforms the data into a model is actually
20 tremendously simple and very principled from a
21 scientific standpoint. So if I had slides, one thing
22 I like to do in forums like this is put up the
23 Wikipedia pseudocode for the so-called back
24 propagation algorithm for neural networks which the
25 previous speaker mentioned. And that pseudocode is

1 literally a simple loop with about ten lines of code
2 in it.
3 And a real working version of it wouldn't
4 be that much more complicated. And it's doing the
5 most obvious thing you can possibly imagine, which
6 is essentially going through the training data and
7 adjusting the parameters or nods of the model in
8 order to minimize some -- you know, usually accuracy
9 or error-based cost function, okay? So that
10 algorithm is not opaque at all. It's entirely
11 transparent.
12 Sometimes, you know, when I talk to people
13 who aren’t in the field, they naturally assume that
14 machine learning algorithms -- you know, the code for
15 them might look like something like I imagine the code
16 to a video game like Grand Theft Auto looking, you
17 know, hundreds of thousands of lines of spaghetti code
18 with all these special cases and details, and it's not
19 like that.
20 So, then, the natural question to ask next
21 is if the complexity doesn't lie in the algorithms
22 themselves then where does the complexity creep in?
23 And, of course, it's from the interaction of the data
24 being processed to produce a model mediated by these
25 very, very simple algorithms, okay? And so the

1 problems arise these days not so much from the
2 algorithms themselves, which, again, are very simple
3 and operating on very basic, kind of well-motivated
4 scientific principles, the problem is really when
5 you work in extremely large complicated model spaces,
6 of which, you know, deep learning is just one and
7 perhaps the most recent example, the sort of space of
8 models has a lot of sharp corners in it, as I might
9 put it, which allow to you kind of optimize the thing
10 that you're trying to optimize like minimizing the
11 error on the data at the expense of other things that
12 you didn't explicitly ask for like fairness or
13 privacy.
14 And I think if there’s one kind of important
15 adage to understand about machine learning, it’s that
16 basically modern machine learning will not give you
17 for free anything that you don't ask for and specify.
18 And in general, you shouldn't expect it to avoid
19 things that you don't want that you didn't tell it you
20 didn't want. Okay?
21 And this is, I think, the source of a lot of
22 the kind of violations of social norms and values that
23 we’ve seen by machine learning and AI in recent years.
24 So that's a little bit about what can go wrong. Now
25 let me talk a little bit about -- sorry, that's a

1 little bit about how things can go wrong.
2 And, so, with that background, I want to
3 talk about, well, what are the different things that
4 can go wrong, and, most importantly, what can we do
5 about them from a kind of scientific standpoint. So,
6 you know, the things that can wrong are things that
7 I've mentioned already, which is violations of things
8 like privacy or fairness or interpretability and
9 transparency, or even safety or morality, if you like.
10 You know, the sort of logical extreme of this for
11 those of you who've heard of it is, you know, this
12 sort of parlor game or science fiction thought
13 experiment known as the singularity on which AI, you
14 know, sort of -- AI achieves superhuman intelligence
15 to the point that, you know, for lack of a better
16 term, the robots become our overlords.
17 While that's a fun thing to think about, I
18 don't know many sane people in the machine learning
19 community who actually think that that's our sort of
20 gravest technological risk anytime soon. All you need
21 to do is come and see what AI and machine learning can
22 actually achieve right now and compare it to humans or
23 other biological species and you will be deeply
24 underwhelmed by what we can accomplish so far. But
25 violations of social norms are, like, already with us

1 now today and on a very large scale, whether we are --
2 whether we know it or not or whether we're measuring
3 them properly or not.
4 And, you know, I think it's important to say
5 to this audience that I think I and many of my
6 colleagues, you know, we do believe that better laws
7 and better regulations are possible and should be
8 developed. And I'm sure that that's being worked on
9 and is a necessary activity. But I think my opinion
10 is that that will be woefully inefficient in the
11 algorithmic era to actually keep up with the types of
12 violations of social values that we're seeing because
13 it just -- you know, basically human organizations
14 don't scale, and you can't sort of expect to police
15 the sort of violations I'm talking about with sort of
16 regulatory agencies that are pouring over the
17 decisions made by algorithms on a sort of a human time
18 scale and hope to keep up.
19 So an alternative approach, which I'm a
20 great advocate of and as are a growing number of
21 people who do technical work in these areas is to
22 design better-behaved algorithms in the first place
23 and to actually endogenize various notions of social
24 norms inside of our algorithms and asking that our
25 algorithms -- that the actual code obey some

1 definition of privacy, some definition of fairness,
2 some definition of morality, if you like.
3 And, of course, this leads immediately to
4 two very difficult questions. The first difficult
5 question is, you know, how do you define these things
6 as the last speaker said. How do you define
7 algorithmic fairness, how do you define algorithmic
8 privacy? And, then, if and when you can come up with
9 such a definition, it's going to come at some cost,
10 right.
11 So if I have some notion of fairness in
12 models that are used to provide criminal sentencing
13 guidelines, my asking for fairness from that model by
14 gender or race will come at a cost of accuracy. What
15 I'm saying is like a tautology. If I sort of -- if I
16 ask myself to find the model in some space of models
17 which minimizes the error period, and then I ask to
18 find the model that minimizes the error subject to
19 your favorite definition of fairness, the error can
20 only get worse.
21 And so in a model like -- let's say in
22 a setting like criminal sentencing, this means that
23 a cost to accuracy might mean sort of, you know,
24 hard things to think about. It might mean
25 incarcerating more innocent people, or it might mean

1 letting more guilty people go free. So there will
2 be societal and technical costs to imposing these
3 sorts of constraints on our algorithms, but I think my
4 view and the view of many people in the field is that
5 we have to go down the road, we have to decide
6 algorithms that incorporate these values, we have to
7 talk about what the possible definitions for these
8 values are, and we need to study these tradeoffs
9 between the thing that people optimize for in machine
10 learning, which is accuracy, and the tradeoffs to
11 different social norms.
12 Okay. And so what I want to do with most of
13 my remaining time is just tell you a little bit about
14 the sort of very active research that's going on in
15 the computer science and machine learning and related
16 communities on this scientific agenda, sort of picking
17 definitions for different social values or norms and
18 actually encoding those norms inside of our algorithms
19 and then trying to study what the tradeoffs will be
20 with, you know, things like accuracy and other more
21 standard objectives.
22 So let me first talk about the work that
23 goes on in the area of privacy in machine learning,
24 and not just in machine learning but more generally in
25 kind of data analysis and data science. And I think

1 it's helpful to say just a little bit about the
2 distinction between what I'm thinking of as privacy
3 and a closely related and complementary area, which is
4 that of security and cryptography.
5 So security and cryptography, to a first
6 approximation, is a technology about keeping data
7 locked down. It's about controlling access to data
8 and making sure that people who shouldn't have access
9 to data don't get that access by basically hacking
10 into a system that they shouldn't hack into. And this
11 is largely the domain of security and cryptography,
12 and that's one notion of privacy. That's sort of
13 control of your data and making sure it doesn't get --
14 you know, it doesn't get accessed or stolen by people
15 who shouldn't.
16 Here, I'm talking about something a little
17 bit different and more nuanced but in many ways is
18 equally as pervasive and important as notions of
19 security, which is the fact that, you know, you have
20 all of this data that's being collected by different
21 companies and agencies and other organizations. And
22 you might worry about what -- not just sort of, you
23 know, how -- who's accessing that data but what can be
24 inferred about you from that data that isn't directly
25 in the data itself.

1 So the kind of thing that I'm concerned
2 about here is that if your medical record is used as
3 part of a study to build a predictive model, let’s
4 say, for some disease based on symptoms, and then
5 that model is published or used in the field, could
6 it be that the use of that model or the publication
7 of the model, perhaps combined with other publicly
8 available data sets, actually reveals a great deal
9 about your own personal medical status and record.
10 Okay?
11 And, you know, if you go down the road of
12 thinking about possible technical definitions of this
13 type of privacy, I believe that most of you would
14 eventually come to two kind of, I think, important
15 conclusions or desiderata from any sort of privacy
16 definition for machine learning or data science.
17 One is that, you know, you need to account
18 for the fact that any particular data set that you
19 want to, you know, keep private in some technical
20 sense, will not be the only data set in the world.
21 And, in particular, that data set might be combined
22 with other data sets that you don't know about or
23 didn’t foresee or don’t even exist yet but might exist
24 in the future.
25 And one consequence of this that I will

1 state without proof is that this means that any
2 definition of privacy that it involves anonymization
3 is essentially a flawed definition of privacy, right,
4 because anonymization refers to taking the data set
5 that's in front of you and doing things like
6 eradicating personally identifiable information.
7 But the literature and news is, you know,
8 rife with examples where you anonymized one data set,
9 somebody else anonymized a second data set. Those two
10 data sets were combined and then maybe combined with
11 some publicly available information, and your specific
12 data could be backed out of that. You could be, as we
13 like to say, reidentified, or the data set could be,
14 you know, deanonymized as they say.
15 And, you know, I think many people feel
16 strongly enough about this assertion that there is
17 sort of a saying in the field, which is, you know,
18 anonymized data isn't, meaning that, you know,
19 whatever you think you did to deidentify individual
20 identity in a data set, that can often be undone
21 through the unforeseen combination of that data set
22 with other data sets.
23 The other, I think, sort of axiom for any
24 definition of privacy that's important is that in
25 order to have a definition of privacy that still

1 allows to us do anything useful with data, it's
2 important to isolate, you know, the potential harm
3 that comes to somebody as the result of use of
4 their data in some analysis or model-building
5 exercise versus the harm that might come to them
6 just because data analysis reveals some facts about
7 the world.
8 So, for instance, if you were a smoker in
9 the early 1950s before there was discovered a link
10 between smoking and lung cancer, well, when somebody
11 did data analysis and discovered that there was a
12 strong correlation between lung cancer and smoking and
13 you were a smoker, that fact does harm to you, but it
14 doesn't matter whether your data was used in that
15 analysis or not, right?
16 Researchers were going to discover this fact
17 whether your particular data was used or not, and a
18 harm was done to you by the fact that suddenly it's
19 revealed that smoking is bad for your health and you
20 were a smoker and everybody knows it. But that harm
21 was not done to you specifically as a result of the
22 data analysis using your data or not. You were
23 basically -- you know, this harm was going to be done
24 to you whether your particular medical record went
25 into those studies or not.

1 And so there is a very rich definition of
2 data privacy known as differential privacy that was
3 introduced about 15 years ago and has received a
4 great deal of scientific attention in the interim,
5 and now has a very rich theory and a very rich body
6 of algorithms that basically on the one hand meet
7 this sort of very strong notion of data privacy which
8 has to foresee the possibility of triangulation
9 through the combination of multiple data sets on the
10 one hand but still permits sort of powerful use of
11 data.
12 And so, you know, one kind of pseudo-theorem
13 that I will state to you is that everything that we
14 pretty much know to do today in machine learning we
15 know also know how to do in a differentially private
16 way. And it's just a matter of companies adopting
17 this technology and choosing to, you know, do their
18 machine learning and data analysis in a differentially
19 private way. And we're actually starting to see
20 large-scale deployments.
21 Both Google and Apple use differential
22 privacy in meaningful, large-scale ways in some of
23 their services, and maybe more importantly, the 2020
24 U.S. Census, every single statistic or report that
25 they release as the result of the 2020 Census they are

1 planning to do so under the constraint of differential
2 privacy. And so this is an example, I think, of a
3 very promising kind of case study, right? Of course,
4 people have thought about different definitions of
5 privacy and data analysis and machine learning for a
6 long time. There was a struggle to kind of come up
7 with the right definition. Many of us believe that
8 sort of definitions based on anonymization are
9 fundamentally flawed.
10 But then a better definition came along
11 around 15 years ago. There's been a huge amount of
12 research on this particular definition, and, you know,
13 the good news is that in this particular -- for this
14 particular definition of privacy and this particular
15 social norm, it is possible to sort of give these very
16 powerful guarantees at not too great a cost to
17 accuracy or computational efficiency and the like. We
18 can sort of, you know, have the best of both worlds,
19 if you like.
20 So let me say a few words about research in
21 fairness in machine learning and algorithmic decision-
22 making, which is much more recent. It's a much more
23 nascent field than the study of privacy and machine
24 learning and AI, but we already know a fair amount
25 about it. And one of the things we already know about

1 it is that it's going to be a little bit messier then
2 privacy. So my claim is that if you waded into these
3 literatures and you looked at the work that’s gone on
4 in deferential privacy and looked in particular at the
5 definition of differential privacy, you perhaps, like
6 many people, might sort of agree that this is sort of
7 the right definition of privacy.
8 So we already know that there’s not going to
9 be a right definition of fairness in algorithmic
10 decision-making and machine learning. And what do I
11 mean by we know there’s not going to be a right
12 definition? So there’s already from the last several
13 years several examples, several papers which have
14 results of the following form. They basically say,
15 well, you know, whatever your definition of fairness
16 is, wouldn't we all agree you’d want it to have at
17 least the following three properties. And you kind of
18 look at those three properties and you’d say yes, yes,
19 I would definitely want any definition of fairness to
20 at least meet those three properties and probably
21 other stuff, too.
22 And, then, of course, the punch line of
23 these papers is, well, guess what, here’s a theorem
24 proving to you that you cannot simultaneously
25 achieve all three of those properties in any

1 definition of fairness. Okay, so those of you that
2 are -- have kind of an economics or social choice
3 background might know about, like, arrows and
4 possibility theorems for sort of voting systems.
5 These have a very similar flavor.
6 And these -- and so this has very concrete
7 consequences. So in particular, a typical notion of
8 fairness in machine learning would ask for the
9 approximate equality of false positive or false
10 negative rates across different groups. So let me
11 give an example. Suppose you're trying -- you know,
12 suppose you’re a mortgage company and you’re trying to
13 build a statistical model that tries to predict from
14 people’s loan applications and credit history whether
15 they will repay a loan if you give it to them or not,
16 okay? A very natural thing to want to do. And you
17 want this model so that when people apply you can make
18 a prediction about whether they’ll repay or not, and
19 then you want to give loans to people that will repay
20 you and not give loans to people that you predict
21 won't repay you.
22 But because this is a statistical model,
23 you’re going to make mistakes. You’re going to have
24 both false positives and false negatives, right? And
25 we might think of false negatives as really the case

1 that causes harm to the consumer in question, right?
2 False negative being somebody who's creditworthy would
3 have repaid the loan if you didn't give it to them,
4 but you decided to reject them, okay?
5 So we might think of false negatives as a
6 harm inflicted on a consumer, and false positives are
7 sort of the people that got lucky. So a typical
8 definition of fairness would basically say that, look,
9 you're going to make false -- you're going to make
10 false rejections. We're not going to try to prevent
11 that, but across different racial groups, it cannot be
12 the case that the false rejections rates differ
13 wildly. It cannot be the case that the rate at which
14 you falsely reject qualified African American
15 applicants is three times the rate at which you
16 falsely reject qualified white applicants, okay? So
17 this is a very natural constraint. And these
18 impossibility theorems basically say if you ask for
19 that and you also ask for a quality of false
20 positives, i.e., the people got lucky, and one other
21 related condition, you cannot simultaneously achieve
22 all of these.
23 So what this means is that we already know
24 that in fairness we're going to have to simultaneously
25 entertain multiple competing definitions of fairness.

1 And so not only will there be sort of tradeoffs in
2 competition between fairness and accuracy, there is
3 even going to be competition between different notions
4 of fairness. If you optimize for one notion of
5 fairness or constrain for one notion of fairness, you
6 might be damaging or losing on another notion of
7 fairness, okay?
8 But nevertheless, you know, we know this and
9 we have to proceed anyway, and so there's been a great
10 deal of research in the last several years on
11 algorithmic fairness and on different notions of
12 fairness and what the tradeoffs between that
13 particular notion of fairness and accuracy is. And
14 this is an area where, you know, to again echo
15 something the previous speaker said, when you sort of
16 talk about the potential interfaces between technical
17 people and policymakers and other stakeholders, I
18 think there are very, very good starting points.
19 So one thing you can do, for instance, is if
20 you pick a particular definition of fairness like
21 approximate equality of false rejections in a lending
22 application, and you have a data set in front of you,
23 a historical data set of people who did and didn't
24 repay loans, you can trace out empirically -- I can
25 give you -- I would have shown this slide if I'd met

1 the deadline -- I can actually show you an empirical
2 tradeoff where on the X axis would be the error -- the
3 predictive error of your model. On the Y axis would
4 be a numerical measure of the extent to which you
5 violated this fairness notion, so 5 percent would mean
6 -- sort of 5 percent unfairness means that let's say
7 between African Americans and other races there's as
8 much as a 5 percent disparity in the false rejection
9 rates. And I can just trace out a curve for you that
10 shows you the menu of choices you have.
11 So you can get the lowest error, but, you
12 know, if you sort of ignore fairness entirely, that
13 will give you the lowest error but it will give you
14 the worst unfairness. At the other extreme, I can
15 demand that the false rejection rates differ by 0
16 percent across populations. It's a very strong
17 constraint. And I will get the worst error but the
18 most fairness, and in between you'll get something in
19 between.
20 And I think this is the type of, you know,
21 sort of quantification of the tensions that we face as
22 a society in making these kinds of decisions that's
23 the right at least initial interface between, you
24 know, people like me and people like you for lack of a
25 better term, right, because it sort of really shows

1 starkly the choices that you have available.
2 Just to say a little more about fairness,
3 most definitions of fairness, like the ones I've been
4 discussing, actually only hold at the group or
5 aggregate level. So you're only making promises to
6 sort of groups of people in a statistical sense, and
7 you're not making promises to individuals, so, you
8 know, sort of more prosaically, if you are a -- you
9 know, if you're a person of a particular race that was
10 falsely rejected for a loan, you would have repaid
11 that loan, the consolation that you have in these
12 types of definitions is, like, well, we're also
13 rejecting people from other races who would have
14 repaid their loans at the same rate that we're
15 rejecting people from your race, which is sort of cold
16 comfort if you're somebody who was mistreated in this
17 way.
18 And so a lot of recent research, including
19 some of my own, is it trying to move towards
20 definitions of fairness and studying their algorithmic
21 implications that try to make finer-grained promises?
22 Maybe not all the way down at the individual level,
23 but to much finer-grained groups than just things
24 like, you know, race -- you know, top-level race or
25 gender or the like.

1 So these are two social values -- privacy
2 and fairness -- on which in relative terms we know
3 quite a bit already scientifically. And I think we're
4 well on the way to kind of developing both a science
5 and an engineering of designing better algorithms and
6 understanding what the tradeoffs are between accuracy
7 and the various definitions of the social values that
8 we've come up with.
9 My former grad student and colleague,
10 Jen Wortman Vaughan, is giving the keynote tomorrow.
11 And she's done a lot of recent research on
12 interpretability and transparency of machine learning,
13 which is another, of course, important social norm. I
14 think we know a lot less there so far, partly because
15 we just haven't had as much time, but one of the
16 problems with sort of coming up with satisfying
17 definitions of things like transparency and
18 interpretability is that there's fundamentally an
19 observer in kind of the middle of such a phenomenon,
20 right? So when you talk about interpretability, for
21 instance, of a statistical model, you have to talk
22 about interpretability to whom and what reason and in
23 particular the sort of numeracy of the audience you
24 have in mind will matter greatly, right, if we're
25 talking about interpretability to people with like a

1 statistics training, that means one thing. If we're
2 talking interpretability to the average American
3 citizen, well, you know, the average American citizen
4 has not been exposed to linear regression and may find
5 it a little bit bewildering to even talk at all about
6 an abstract mathematical mapping from loan
7 applications to lending decisions.
8 And so I think much of the research that
9 needs to happen on that topic will have to have like a
10 cognitive and behavioral element to it. You'll need
11 to do human subject studies with the type of audience
12 that you're interested in and ask them what they think
13 is interpretable to them or whether you can explain
14 models to them and the like.
15 So I'm almost out of time, but just to sort
16 of quickly touch on a couple of other things that I
17 saw on the agenda, I saw that there was one discussion
18 -- there was one panel title that had a very
19 intriguing name, which was Algorithmic Collusion. And
20 I'm not sure exactly what the context that's meant
21 there is. But, you know, if your concern is that, you
22 know, we might be entering an era where algorithmic
23 decision-making causes in some implicit or explicit
24 kind of large-scale way collusion between different
25 entities, whether it's on things like pricing or

1 decision-making and the like, I definitely think this
2 is already happening.
3 One area in which I'm very familiar with
4 this already is on Wall Street where quantitative
5 trading teams tried to build statistical models to
6 predict the directional movement of stocks and, so to
7 speak, beat the market. And my basic belief there is
8 that there’s a huge amount of implicit sort of
9 collusion going on there, and it's really because, you
10 know, when we all use the same or similar data sets,
11 and when we all use the same or similar algorithms to
12 train our models, then even if we think we're clever
13 and independent and creative, we are going to be
14 strongly correlated just through the data, right?
15 If we’re trying to predict the same thing
16 and we’re using similar data sets and similar methods,
17 then no matter what else we do -- everything else
18 we’ll do is a second-order effect from the fact that
19 the data itself will correlate us. And so I think
20 that this is an interesting topic on which there is
21 probably interesting scientific things to say but I
22 haven't thought about it yet, and I don't know of a
23 large body of research on it
24 But I'm out of time, so let me stop there
25 and let the agenda move on.

1 (Applause.)
2 MS. GOLDMAN: Well, thank you so much,
3 Professor Kearns, for that great overview and
4 introduction to all the issues that will be covered in
5 this hearing.
6 DR. KEARNS: Okay, thank you.
7 MS. GOLDMAN: So now it is 10:15, and we're
8 going to be taking a little break until 10:30, at
9 which time we will be back for the first panel.
10 (End of Presentation.)
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 UNDERSTANDING ALGORITHMS, ARTIFICIAL
2 INTELLIGENCE, AND PREDICTIVE ANALYTICS THROUGH
3 REAL WORLD APPLICATIONS
4 MS. GOLDMAN: So we're now going to begin
5 the first panel on Understanding Algorithms,
6 Artificial Intelligence, and Predictive Analytics
7 through Real World Applications. As I mentioned, in
8 case someone had just come in, I'm Karen Goldman. I'm
9 an attorney advisor in the Office of Policy Planning
10 at the Federal Trade Commission. And this is my
11 comoderator, Dr. Harry Keeling, who is an Associate
12 Professor in the Department of Computer Science here
13 at Howard University.
14 So we hope that this panel, which will cover
15 applications that are currently in use and on the
16 horizon, will provide a sense of the variety of uses
17 to which these digital tools can be put and highlight
18 that no single application is truly representative of
19 their use.
20 I just want to mention that anyone in the
21 audience who would like to ask questions of panelists
22 should write their questions on the notecards that are
23 being passed out and will be collected later on with.
24 With that, I'd like to introduce the
25 distinguished members of this panel. So we have Dana

1 Rao, who is an Executive Vice President and General
2 Counsel of Adobe. Next, we will have Henry Kautz, who
3 is the Division Director for the Division of
4 Information Intelligence Systems in the Directorate
5 for Computer and Information Science and Engineering
6 at the National Science Foundation.
7 Then we will have Angela Granger, who is
8 Vice President of Analytics at Experian. And then
9 Melissa McSherry, who is Senior Vice President, Global
10 Head of Data Products at Visa. We have Michael
11 Abramoff, who is the Founder and CEO of IDx, and
12 Professor of Ophthalmology and Visual Sciences at the
13 University of Iowa and also Professor of Engineering
14 and Computer -- of Electrical and Computer Engineering
15 and Biomedical Engineering.
16 And then we will have Teresa Zayas Caban,
17 who is the Chief Scientist at the Office of the
18 National Coordinator for Health Information
19 Technology.
20 So with that, Dana, would you like to begin
21 your presentation?
22 MR. RAO: Thank you.
23 Thank you. Thanks for being here. So the
24 first thing I wanted to just sort of get out there,
25 I'm a lawyer, and people are like, why are you talking

1 about AI, and I thought I'd put it out there because
2 there are some very distinguished computer scientists
3 on this panel. So I was a actually an electrical
4 engineer undergrad and going to a university, and so
5 when I was at law school, I was going to write a
6 paper, a note for the journal, and this book was on my
7 dad's desk, Understanding Neural Networks. This was
8 back in 1996. So I thought, oh, that would be fun to
9 read, and I read it, and I wrote my paper, which got
10 published, called "Neural Networks -- Here, There and
11 Everywhere," which was a wildly inaccurate
12 characterization of where neural networks were in
13 1996. So don't come to me for your stock advice, but
14 it was -- it's been a fascinating topic for me, and at
15 Adobe, we're really interested in this topic.
16 And for us, AI is special because we have
17 this entire business that's focused on helping people
18 be creative. And creativity is a part of the brain
19 that doesn't follow rules. It's unstructured, and
20 traditional software programming is a very structured
21 form of algorithms. It's predictive. You understand
22 the rules, and you understand how to characterize it,
23 and that's actually not a great fit for creatives who
24 tend to break rules.
25 And so our products have always struggled to

1 bridge that gap between innovation and creativity and
2 the structure form of traditional computer
3 programming. And AI bridges that gap, and it really
4 allows us to create tools that are better for our
5 creative customers.
6 So when we think about how we look at AI and
7 digital creativity, we're really focused on minimizing
8 the mundane, eliminating those repetitive tasks that
9 everybody has in their day. And so for creative
10 professionals, there’s a lot of complexity in the
11 tools and in setting up the camera shots or the video
12 shots that are not actually the highest value add that
13 they have, where they’re really trying to get their
14 artistic sense across or fulfill the goal of a
15 marketing campaign as they create content for it, the
16 complexity of adjusting each pixel’s luminance or the
17 color or the background or the lighting gets in the
18 way of them actually doing the part of the work that
19 they're getting paid to do. So that's where we're
20 really interested in using AI is it would eliminate
21 those mundane tasks.
22 And we also at Adobe, we’ve noticed there’s
23 a huge demand for content now, and that's either
24 because there’s social media channels and people are
25 posting content all the time on Instagram and Snapchat

1 and Facebook, or on ad campaigns -- digital media
2 advertisement campaigns where -- digital marketing
3 campaigns where you are personalizing content for each
4 consumer. So there’s a huge demand for content, more
5 than ever before, and our creative professionals need
6 to be able to create content at a higher velocity, and
7 that's what AI is helping us do.
8 So when we think about AI, we think about it
9 in the creativity space in two buckets -- content
10 understanding, computational creativity. And Adobe
11 also has an experience intelligence business. I’m not
12 going to talk about that much today, but just for
13 transparency, we also have this other business that
14 also uses AI in a different way.
15 Content understanding is really trying to
16 get behind what’s in an image, for example, or a
17 video. So it's easy to look at an image of a cat and
18 say there’s a cat, or there’s a house and just do sort
19 of basic object recognition. What AI allows you to do
20 is provide that insight into the image and add an
21 abstract layer, a conceptual layer above what you
22 typically can do pre-AI so we can understand things
23 like actions and concepts and styles and sentiments,
24 so just abstract concepts that are in your image that
25 the AI can infer from looking at it.

1 And so we have a couple of demos that we're
2 going to show. We're hopeful they're all going to
3 work correctly. I think this is going -- yeah, it's
4 going. And these -- this deck will be published in
5 the Adobe public policy blog, so anyone who wants to
6 see the full deck and watch the videos through can do
7 that. But we're just going to talk through a few -- a
8 couple seconds of these.
9 So this is a person in the, let me just go
10 back here. Set this up. So this is a person using
11 our stock photography service. And so they wanted to
12 start a creation. And so they wanted to be able to
13 say, I have an ad campaign for Nike, how should I
14 start. And they go to our stock photography site
15 and they just search for things to sort of -- as
16 inspiration for the ad campaign.
17 And so for example, in this example, this
18 person's going to say, you know, I see this image of
19 this woman with a ribbon jumping. That sort of
20 captures the aesthetic of what I want. And here we
21 go, here we go. And so she -- say they choose this
22 picture, and then what Adobe Stock does, it recommends
23 other pictures that are very similar to this picture.
24 So in this case, she says, okay, I like this, this is
25 a good start for me, and then Adobe Stock at the top

1 does sort of a normal picture recommendation. Here
2 are other pictures of people with ribbons, and that
3 may be what you're looking for.
4 But in this case, that's not what we want.
5 Like, Nike actually wants this sort of freedom. And
6 so we select the woman jumping, and our AI understands
7 that what we want is actually the action of jump.
8 Like that’s what we want out of this picture, not the
9 color, not the ribbon, not the blue sky. We want the
10 action of jump. And so now we actually recommend
11 pictures that are about jumping.
12 So we can take the concept of that picture
13 and using AI understand, okay, they actually wanted
14 jumping, and so now we can just show them these other
15 pictures.
16 Now, the next level is we say, okay, well,
17 Nike didn't really want a picture of random people
18 jumping. It was actually supposed to be a family
19 picture. So we take family and we use the jump
20 concept from the first image, so you see how they’re
21 stacked on the right, and now you have families
22 jumping. And now the creative professional could say,
23 that’s where I want to start, I want to choose one of
24 those pictures and start my campaign from there.
25 So how do we do it? So what we do is our AI

1 will analyze these -- in this case, an image -- and
2 look for the concepts behind it. So you can see in
3 the middle, there's concepts, and on the right, there
4 are percentages. The percentages are the confidence
5 that our AI is actually accurately predicting what is
6 going on in there. But what you can see is we've
7 analyzed those faces and we've analyzed the context of
8 the picture, and you can see that where you said, oh,
9 there's happiness there, there's love there, there's
10 joy there, we've understood the abstract concept of
11 those pictures. And so you can go, if you're a
12 creative professional, and say I need pictures, my
13 theme is love, you can type in love as a search term,
14 and you're going to get a wide variety of images, but
15 they're going to have this concept in them.
16 You can also look for families, right? And
17 it will understand that the connection of these three
18 people plus the expression on their face means that
19 they're a family. And you can understand -- and you
20 can search for concepts like family as part of this.
21 And so you can see all the different kinds of
22 categories that you are able to search on using our
23 Adobe AI technology to analyze what is actually going
24 on inside the picture.
25 We also have a PDF and Acrobat service, you

1 know, and that has lots of text, and we've actually
2 run our AI on the text to understand the intelligence
3 behind the words. And we have married that up to
4 images to allow you to do automatic phrasing. And,
5 again, we can do very basic captioning. So you put
6 your photo there, and we can say couple on a bike and
7 that's object recognition. But then we see AI, then
8 there's a little slider you can see that's moving.
9 And you can say I want to see what the AI thinks this
10 is. And it says young couple on a bike, or in this
11 case, it said beautiful peacock, right? So it
12 understands not just the image but also the concepts
13 behind the image. So if you wanted to search for
14 "beautiful," you'd get that peacock, for example.
15 So these are the techniques that are being
16 used when we talk about content understanding, the
17 first part of how we looked at AI and creativity. You
18 know, traditional machine learning, it's traditional
19 deep learning, and when you look at all these things
20 like aesthetics and style and color, as part of the --
21 we train our AI to understand these concepts, and then
22 we are able to provide these services to our creative
23 professional.
24 The second piece of what we do is try to
25 make the creative professional day's faster. And

1 that’s what we call computational creativity. And
2 that is trying to help their work flow. How do we
3 help them do those tasks even faster than they used to
4 have to do under traditional software? So here’s an
5 example. Let's say somebody wants -- Macy's wants an
6 ad campaign and they told you to go out and shoot a
7 cityscape at night, and you go out and you spend six
8 months getting this shot. It had the right lighting,
9 the right building, the right angle, and you’re like,
10 all right, I'm great, I’m happy.
11 And then you turned it in, and Macy's was
12 like, you know what, we’ve changed our mind, we want a
13 different setting. We want it to be the sunset. And
14 so then traditionally, you’d have to go spend another
15 six months reshooting this picture trying to get the
16 lighting correct.
17 So with our AI, we can automatically segment
18 out the part of the picture that’s of interest to you,
19 and then that's the cityscape. And then we let you
20 import another picture that is of the desired lighting
21 and sky that you want. And with one click, you can
22 now take that lighting and put it in your picture.
23 So that's probably not 100 percent of what
24 the creative professional wants for their Macy’s
25 campaign, but it's probably 80 percent or 90 percent

1 of what they want, and now they can take this picture
2 and make it into exactly what they want with very
3 little extra effort. So you've just taken six months
4 of extra work, of not exciting work, that was not the
5 fun part of their day. The fun part of their day was
6 setting up that shot to get that image in the first
7 place. And now they can take this and they can go
8 back to Macy's, and if they come back and Macy's says,
9 you know what, we've changed our mind, snowy, blue-sky
10 day, five minutes later, you can just change. And so
11 the AI really helps drive that routine out of your
12 day.
13 Another example is what we call neural
14 stylization. And so, again, this is the idea that
15 we've been able to understand the style of an image.
16 And so we've trained our AI demonstration on the style
17 of different famous paintings. And so if you have
18 your photograph on the left and you said, I want it to
19 look something like the interpretation of these two
20 different paintings, you can do it. All it does is
21 understand the style of whatever painting you put in,
22 and it's just the style of it. So it's not just
23 copying the colors broadly like you might have
24 expected pre-AI. It understands what the style of the
25 image was and applies it to the image.

1 So just understanding that concept of -- I
2 think this is going to play. And so this is not just
3 for creative professionals. This is for hobbyists.
4 You can take your own pictures and you can upload
5 whatever artist you want and it's going to take the
6 style of the artist and apply it to your picture. And
7 it understands that concept.
8 We can also use AI and we do use AI for our
9 video editing products. So this is a project called
10 Cloak, and this is a normal example where you have --
11 you shot a scene and then in post-production, you want
12 to get rid of something you don't like. In this case
13 you don't want that couple there. So in using AI, we
14 segment the image and understand who's in the image
15 and who they are, and we can also fill in the
16 background with copied pixels to make the background
17 look perfect.
18 So on the left is the original footage, and
19 on the right is post-AI, and it looks like they've
20 just vanished, right? And then that used to take
21 months of work to do to edit two people walking out of
22 the video, and now you can do it in minutes.
23 So as I mentioned, we also have an
24 experience intelligence business. This is the other
25 side of our business. This is a digital marketing

1 business that allows you to target advertisements and
2 allows chief marketing officers to understand what the
3 content in the campaign is doing. So we provide that
4 service and use AI there as well. We use it to help
5 you predict the results of a campaign before you even
6 launch it. We may say this is going to be successful
7 in the northeast, or this is going to be successful in
8 California based on our analysis of customer data from
9 interacting with their website. That's another way
10 we're using AI at Adobe.
11 So I think the question is how we get there.
12 How do you actually produce the AI, and I know there's
13 going to be a lot of people talking about the nuts and
14 bolts of the computer science so I'm not going to
15 spend too much time on this, but this is how Adobe
16 does it. Our AI product called Sensei. And this is
17 the architecture.
18 And so what we do is -- what we do typical
19 of any neural network, we have the neural network and
20 then we train it with data, and we train it for an
21 outcome. And using this architecture, we're able to
22 create the neural network; we freeze it in place; and
23 we ship it PhotoShop; we ship in Premiere, and that's
24 the result you see as a consumer.
25 So the principles -- this is my second-to-

1 last slide -- the key principles for training AI that
2 is important to Adobe and just a takeaway for everyone
3 is how do we make this product work as well is we need
4 millions of pieces of data to train it. You need lots
5 of examples of artists; you need lots of examples of
6 images in order to train a neural network to
7 understand the insights that we're able to show you.
8 So when you think about how do we make AI
9 beneficial, how do we get the rewards of AI, you need
10 access to data, you need access to a lot of data and
11 you need access to a variety of data, and that variety
12 of data will make your neural network accurate. And a
13 variety of data will also eliminate bias.
14 You can imagine bias when you're looking for
15 images, that is inherent because you may have trained
16 your AI on that particular kind of a person, and if
17 you go searching for a job or an occupation, you're
18 always going to get that person because that's what
19 you trained it with. So the wider variety of data you
20 put into the AI, the more likely it is your results
21 are going to be unbiased.
22 So thank you for your time. This is our
23 presentation. Creativity in AI, with AI is what Adobe
24 is focused on. It's how we believe AI will help
25 transform the creative professionals for today and

1 tomorrow. Thank you.
2 MS. GOLDMAN: Thank you so much for that
3 colorful and creative presentation.
4 So, next, Henry Kautz will begin his
5 presentation.
6 MR. KAUTZ: Thank you. So I'm going to
7 focus my talk on the work we've been doing at NSF to
8 support AI applications for social good. So when we
9 look at a proposal, we have two major criteria.
10 First, we want to advance science or engineering
11 looking at fundamental advances, but we also consider
12 potential for broader positive impacts on society.
13 Now, the traditional broader impacts that
14 were frequently mentioned in proposals, we're training
15 graduate students and potential future applications of
16 the result. So someone, say, I'm doing this
17 fundamental research, and maybe someone in the future
18 will come along and do something to benefit society
19 with us.
20 But, increasingly, we see that the
21 fundamental science and these broader impacts are
22 entwined, that as you work on an application for
23 social good you discover new questions that require
24 fundamental scientific advances. And from those
25 advances, you discover that there are new

1 opportunities.
2 So AI and broader impacts. So AI methods,
3 taken broadly, that includes machine learning,
4 knowledge representation and reasoning, and what we
5 might call deliberative intelligence, making optimal
6 decisions, are being used by researchers in every
7 discipline that's funded by NSF. I'm from the
8 computer science, and my particular division funds a
9 lot of the fundamental work in AI, but there's really
10 no area of NSF now, including the social sciences,
11 where you don't see people talking about AI. And,
12 increasingly, we're partnering with other agencies,
13 funding or taking advantage of work in these fields.
14 So we've seen -- over the last decade, we've
15 grown up quite a rich portfolio of what we call cross-
16 cutting programs. So these are interdisciplinary
17 funding opportunities that involve multiple
18 directorates within NSF and sometimes with other
19 agencies. Some of the most important are the Smart
20 and Connected Health Program that we run with NIH.
21 And so there, we are looking at AI research that is a
22 bit more applied than traditional work funded by NSF
23 but is not yet ready for the kinds of actual clinical
24 uses that NIH would fund. So we both put money in
25 there, and then we help bridge the gap between those

1 agencies.
2 Smart and Connected Communities looks at
3 applications of AI to all kinds of problems facing
4 urban life from pollution, policing, and violence,
5 transportation, other issues. We’ve had a program for
6 several years now called Big Data in science and
7 engineering, and that is to support broad
8 collaborations -- or collaborations that can cover a
9 number of fields. So you might have material
10 scientists together with a computer scientist or, you
11 know, electrical engineer together with the computer
12 scientist or even medical people.
13 And through that Big Data program, we’ve
14 also funded what are called big data hubs, so the idea
15 that these are a set of universities that act as
16 resources to all of the universities in that region
17 for activities such as helping making connections to
18 government agencies. And through that, we’ve had
19 programs like the Civic Innovation Challenge.
20 One of our most recent programs that is
21 particularly relevant for broader impacts is one
22 called the Future of Work at the Human Technology
23 Frontier. And it's a very interesting combination of
24 directorates -- computer science, engineering,
25 education, and then our social, behavioral and

1 economic sciences.
2 So we're now looking at the future of the
3 workplace and in particular how AI will be impacting
4 that future. So we want to fund both the kind of
5 technology we might see in the future. So, for
6 example, in a recent -- we just completed the first
7 year of the program, and one of the awards was on
8 smart classrooms, so how we might integrate AI as a
9 teacher's assistant, and not replacing a teacher but
10 assisting a teacher.
11 But we also will be looking for work where
12 technologists work with social scientists to look at
13 both the positive and the negative consequences. Will
14 AI throw millions of people out of work? That's
15 absolutely an open question. And if you look back at
16 the history of science and technology, you can make
17 quite good arguments either way that AI will lead to
18 permanent unemployment or that AI will lead to new
19 opportunities for employment.
20 This is another example of the work from
21 this most recent program -- solicitation. So Whole-
22 body Exoskeletons for Advanced Vocational Enhancement.
23 So, here, we're looking, you know, at something a
24 little bit different than your traditional robotics
25 for manufacturing but augmenting the human worker to

1 give the human worker superhuman strength and
2 endurance, or as I mentioned in classroom teaching,
3 where a system that is monitoring a classroom and
4 noticing when students -- those students who have
5 become apparently disengaged are not working or not
6 attending and can perform such tasks as simply
7 alerting the teacher or engaging in a personalized
8 activity with the student.
9 So one of our very largest grant programs
10 is called Expeditions in Computing. These are
11 typically $10 million over four to five years. So,
12 here, we're really looking for research of the highest
13 intellectual merit. All of our reviewing is a system
14 called peer reviewing, where we get unbiased
15 scientific experts from the community to review. And
16 in Expeditions, we have multiple layers of viewing
17 because we really want to get the best of the best.
18 And in addition, these -- the work we fund
19 should address the nation's greatest needs. So to
20 give just a case study of the synergy between positive
21 broader impacts and scientific merit, I'd like to just
22 mention some of the work going on at the Institute for
23 Computational Sustainability, which is a -- the result
24 of actually two successful Expeditions in Computing
25 that went to a consortium of Cornell, Stanford, and

1 University of Southern California.
2 So the problem here is looking at
3 sustainability problems, and by sustainability, we're
4 looking at environmental sustainability, economic
5 sustainability, resources, social sustainability, very
6 broadly, as complex problems that are really too
7 difficult to solve with human intelligence alone. So
8 we want to employ AI techniques and large amounts of
9 data to solve optimization -- essentially resource
10 optimization problems that are far beyond the kinds of
11 linear optimization that most of the people in this
12 audience would be familiar with.
13 These are highly nonlinear problems where we
14 must model uncertainty. So we can't -- we just can't
15 ignore the fact that many -- there are many variables
16 that are unobserved. Okay.
17 Now, you might think that, well, these are
18 all different problems, but what has been so
19 fascinating by this Expeditions is that problems that
20 seem to be quite different often have very -- have
21 shared technical solutions, okay? So this is a subway
22 map that the research group created. And as we see,
23 each of the tracks of the subway, the six tracks --
24 the six tracks are scientific themes. So pattern
25 decomposition, crowdsourcing, mechanism design, so

1 social choice theory, and economics, spacio-temporal
2 modeling probabilistic inference, and sequential
3 decision-making. And then each of those tracks is
4 going through the stops, where the stops are the
5 particular application.
6 So in each application you had domain
7 experts. So let's say there's one there on landscape
8 scale conversation and rural communities. That
9 included, you know, people who knew a lot about that
10 topic and had been studying and working with
11 communities in Ecuador, but it made use there of
12 temporal modeling, probabilistic inference, and
13 sequential decision-making. So you see it's quite a
14 variety here -- flight call detection, and I'll
15 mention that again, wind and solar forecasting, all
16 the way over to microbial fuel cells.
17 Now, but one thing I should point out is AI
18 covers many things. There's sometimes a tendency
19 because of the great success of what are called
20 artificial neural networks to say that that is AI.
21 And as we just saw from the previous speaker,
22 artificial neural networks are wonderful when you're
23 dealing with patterns, doing pattern recognition, and
24 essentially trying to emulate those parts of
25 intelligence that don't involve essentially logical

1 thinking but are more based on pattern recognition and
2 intuition, the kinds of problems we don't think about
3 when we solve them -- recognizing your friend's face,
4 right? We don't think consciously about it.
5 By and large, the work in this particular
6 set of projects, though, involves what we may call
7 your Type 2 intelligence, your deliberative rational
8 intelligence where you have many alternatives to
9 consider. In fact, there is such a large number of
10 alternatives, you can't simply enumerate them all one
11 after the other. You have to have very clever ways of
12 essentially searching through an enormous, sometimes
13 infinite space of possibilities and narrowing in on
14 those points that are near optimal.
15 So just going down a little bit deeper, the
16 problem of data -- of decomposition in big data. So
17 this is -- so a core technical problem. You have some
18 kind of very complex signal, and you want to reduce it
19 to something simpler, right, to a small -- the one
20 measurement or a small number of measurements. So
21 this is also called dimensionality reduction, source
22 separation, sometimes called segmentation. But it
23 makes use of a body of algorithms that have come up in
24 computer science, electrical engineering, and
25 particularly more and more in work in AI.

1 So we had a -- there were a series of
2 projects, one on detecting gunshots. And you can
3 imagine security applications in a city. Another one
4 detecting elephant calls. So you can put out audio
5 monitors in the jungle and use that to conduct a
6 census of elephants, right, based on their calls.
7 That same work was then used to detect birdcalls of
8 actually birds in flight for a project with the School
9 of Ornithology at Cornell. And perhaps, surprisingly,
10 is with very few changes, that same algorithm was used
11 in a project on crystal phase mapping, which is in
12 material discovery, so a problem where you're coming
13 up with a mix of new materials, you hope they have
14 some property, and you're analyzing the results of
15 shooting x-rays at those new materials.
16 Another example -- my last example here --
17 is dealing with hydropower in the Amazon Basin. So
18 there are a great potential for getting more
19 hydropower from the Amazon Basin. And, in fact 170
20 dams have already been built or under construction,
21 about 300 dams are planned or proposed.
22 Now, there's obviously a big problem
23 here. If all of these damns are built, not only will
24 there be quite a lot of devastation to wildlife, but
25 they will become less effective because one damn is

1 going to affect the water flow to another damn.
2 So you have to look at this as a multi-objective
3 optimization problem to balance off energy,
4 fisheries, transportation, and navigation. Obviously,
5 as you put in more dams, you're going to make river
6 transportation more expensive, and finally looking at
7 the long-term effects, how will all these damns affect
8 the natural flow of sediment and nutrients and how
9 that affect farming. So this becomes a multi-
10 objective optimization problem.
11 And then the goal is to look at the
12 tradeoffs between these different factors and have a
13 new algorithm that can present, well, here is the
14 possible best tradeoffs. There's no single best
15 tradeoff, but you can look at that any solutions that
16 don't fall along this line are provably worse, so
17 they're worse in some respect and no better in any
18 other respect. So this tremendously reduces this sort
19 of infinite space of the number of dams and the
20 placement of dams to one that now can be decided by
21 humans. Yeah, that's showing where they're the dams.
22 And interesting that this same effort has
23 led to startups. For example, ATLAS AI, that is
24 basically a for-profit AI for social good company.
25 This also received funding from the Rockefeller

1 Foundation, looking at providing -- helping developing
2 nations be more sustainable in their agricultural
3 practices. Networks of CompSustNet, a larger network
4 that includes this group of these three universities
5 with others to address these problems.
6 And with that, I'll conclude. Thank you.
7 MS. GOLDMAN: And thank you so much for
8 showing us the diverse portfolio that NSF is
9 supporting.
10 And, now, Angela Granger will begin her
11 presentation.
12 MS. GRANGER: Thanks. Sorry, it's a little
13 tight up here, so we thought that would be the better
14 route to get around.
15 I lead analytics for Experian, and one of
16 those areas that I'm responsible for is credit scoring
17 product development, and for those of you that don't
18 know, Experian is a global leader in consumer and
19 business credit reporting and marketing services. We
20 support clients in over 80 countries, and we have
21 approximately 17,000 people in 37 different countries.
22 We believe it's our responsibility to assist
23 lenders in managing consumer credit risk and
24 empowering consumers to understand and responsibly use
25 credit in their financial lives. We're committed to

1 being the consumers' credit bureau, and I thank you
2 guys for having me here today.
3 To set the context for today, there's a lot
4 of different areas of application for credit scoring,
5 so we're going to -- I'm going to specifically talk to
6 scores used to assess eligibility for credit where
7 adverse action may be taken. The example was used a
8 couple of times earlier today specifically of
9 application of credit for an example where you could
10 be approved or declined, your application for credit.
11 That would be the credit scoring context we're talking
12 about today.
13 Benefits of AI or machine learning, for both
14 lenders and consumers in our industry, are ultimately
15 better lending decisions. If you have greater
16 insights into the data that you're using, better
17 accuracy in the scores, you're going to have better
18 decisions being made.
19 And, secondarily, financial inclusion.
20 Where we're really finding the power of AI and machine
21 learning techniques is our ability to evaluate new
22 data sources more quickly and incorporate that new
23 data into credit scores, thus broadening the access
24 for credit for people who maybe have thin credit or
25 are new to credit and don't have a credit file with us

1 today.
2 Where we like to start is with the data. If
3 you think about predictive modeling, and any kind of
4 modeling for that matter, it's important to understand
5 the data that’s feeding into the model. For us, we
6 talk about traditional credit data. And when you
7 think about traditional credit data, what we refer to
8 is what you typically find on the core credit
9 databases at the major credit reporting agencies. And
10 this includes information around what we call trade
11 lines or account-level information where you get
12 access to a consumer’s payment history on a certain
13 type of account, their outstanding balances, that sort
14 of thing.
15 We also have information on inquiries that
16 are made into the credit bureau for applications for
17 credit for example. And we have public record
18 information, particularly on bankruptcies. We also
19 maintain some additional information that you might
20 think of as being part of a credit application, such
21 as income and employment.
22 We also like to talk about alternative
23 credit data. So this goes by many terms. In our
24 industry, when we say “alternative credit data,” we
25 really mean data that is not on that core credit

1 database that I talked about a minute ago. So types
2 of alternative credit data that aren't reported to the
3 core credit database today can include rental
4 payments, asset ownership, alternative financing such
5 as payday loans, short-term loans, rent-to-own-type
6 loans.
7 There's additional public record information
8 out there that's not on the core credit database.
9 And, most recently, we've incorporated consumer
10 permission data.
11 Both alternative data and traditional credit
12 data have been found to be very predictive of a
13 consumer's creditworthiness. And, particularly, the
14 alternative data comes into play in those cases of
15 thin file and no-hit-type consumers that I mentioned a
16 minute ago.
17 The Fair Credit Reporting Act regulates the
18 collection, dissemination, and use of consumer credit
19 information, and so all data used in credit scores are
20 what we would call FCRA-complaint. What does that
21 mean? That means the data needs to be accurate, so
22 the credit reporting agencies must do their best to
23 ensure their data is accurate. The data is
24 disclosable, so consumers can see that information.
25 Consumers can get one free credit report every 12

1 months, and they can see their credit information if
2 they're denied credit as an example.
3 The data furnishers also play a role in the
4 process. They have to confirm information when
5 disputes happen, and they're held to certain
6 turnaround times as well as part of the dispute
7 process. And, lastly, we were set up pretty nice
8 earlier around fairness. Fairness is another part of
9 the FCRA. So scores are -- they cannot discriminate
10 based on these different ECOA factors such as gender,
11 marital status, race, and religion.
12 So for about 30 years, we've been using
13 scores kind of in their current form, which means
14 they're using this core credit information that I
15 talked about earlier. And so between that and our
16 experience over time, we've come up with things that
17 are generally acceptable in our space, data that
18 complies with those FCRA rules that I mentioned
19 earlier, proven payment information, rental data,
20 account transactions from your demand deposit accounts
21 are generally deemed acceptable. Generally not
22 acceptable are things like social media data, you
23 know, who your Facebook friends are sort of things,
24 and any data that could discriminate in decisions or
25 that could be discriminatory, I should say.

1 Under consideration right now, we're looking
2 at education level, again, something to help us in
3 that new-to-credit space. Think of students
4 graduating from universities and having that
5 information available so that they can more easily get
6 credit and join the credit ecosystem.
7 So one of the things about our industry is
8 not only is the data itself, which we just went
9 through, regulated but the scores or the models are
10 regulated as well. There’s regulatory guidelines
11 around accuracy and fairness that have been put out by
12 the OCC. Those documents or those guidelines, I
13 should say, are pretty extensive. They cover the
14 model development process, they cover model use, they
15 cover model monitoring, when to redevelop. And they
16 create quite an extensive amount of documentation.
17 And in order to meet these model governance
18 guidelines, many of our clients -- so think of, you
19 know, big banks, big lenders -- have had to create
20 entire staffs just to take on this model governance
21 requirements.
22 We talked about the controls around
23 discrimination which lead to the need for
24 transparency. And then in the FCRA, we are also
25 required to provide your top four reasons for your

1 score being what it is as well. And so the need for
2 transparency, or what we call explainability in
3 scores, is very high.
4 Some key considerations when developing
5 credit scores to meet all these needs, I won't go
6 through all of these in particular, but they really
7 cover the full life cycle. We talked about, at one of
8 the earlier sessions, generalization. So our models
9 need to essentially replicate. They can't just work
10 really well on the training sample. They have to work
11 well in production. If you think about credit scores
12 in use today -- think about mortgage in particular --
13 the credit scores being used there are about 20 years
14 old, right? So these models need to continue to
15 replicate and still rank-order consumers in terms of
16 their creditworthiness.
17 Today, models have an average shelf life of
18 about three years, so we're looking at AI to help us
19 get models to market faster. Some research that we
20 did, we tested several different techniques around
21 machine learning. I won't go into each of them. You
22 can see that here. But suffice it to say the gradient
23 boosting models are the ones for credit scoring and
24 credit risk in particular that seem to be rising to
25 the top.

1 When we let the machine run by itself, these
2 are the type of results we get. We see anywhere
3 between a 5 percent to 10 percent lift depending on
4 the situation. This is a more generic sample for auto
5 and bank card, so we see about a 5 percent lift if you
6 were to do the math here. But our clients report
7 anywhere up to a 15 percent lift as they start to
8 really look at specific portfolios or specific
9 lenders.
10 This, however, is when you just let the
11 machine run itself and you don't take into
12 consideration some of those things we talked about
13 earlier.
14 We do something that we call model
15 refinement, and this is where you have to go in and
16 ensure your model is working as expected, that you can
17 explain what’s happening. You want to make sure that
18 a credit score doesn't go down if a consumer has made
19 some impact to their credit such as paying off some of
20 their debt or lowering their utilization. And if you
21 don't do this refinement and you don’t understand
22 what’s happening under the covers, that can happen.
23 So when you go in and you refine the model
24 through the requirements that we talked about before,
25 you’ll see that the lift in performance from the -- in

1 this case, extreme gradient boosting methodology, is
2 lessened. So in our particular example, the lift went
3 from 5 percent to 2 percent. In other examples, we've
4 see that 15 percent or 10 percent lift come down to 5
5 to 8 percent, right? So on average, we're seeing
6 about a 5 percent lift in accuracy from applying some
7 of these techniques outside of our traditional
8 regression methods.
9 This is just another example of addressing
10 overfitting, which tends to be a problem with some of
11 these new methodologies that aren't -- haven't been
12 used in practice as long. What you tend to do if you
13 throw all of the data into the machine and let it do
14 its work, we have over 2,000 attributes, variables,
15 characteristics that we will throw into a model, and
16 it will use almost all of them if it can, right.
17 And that tends to overfit and the model
18 doesn't generalize. And so you do have to go in and
19 manually intervene and not let the machine do all the
20 work.
21 Some of the advantages for AI in credit
22 scoring go beyond just the modeling. You know, I
23 mentioned a 5 percent improvement, and I'm sure you
24 guys are all sitting there, going, whoo, 5 percent, 5
25 percent, right? But in the credit risk world and

1 creditworthiness world, we have very predictive models
2 today. And so a 5 percent improvement is actually a
3 big improvement. The data that we use in the models
4 is very accurate, and so we get very good models. So
5 5 percent improvement is significant, but we're
6 looking to use machine learning and AI methodologies
7 across the model development life cycle and not just
8 in the model development itself.
9 Probably most importantly to take away from
10 today is in credit scoring. Credit scores are static
11 models. So most of us when we think of AI think of
12 realtime updating, self-learning type models. Those
13 are not in use in our industry today. These are
14 static models. So while we're looking at these
15 additional techniques outside of regression, we're
16 still talking about static models. I mentioned the
17 turnaround time or the shelf life of a model is about
18 three years right now. With these new techniques,
19 that's going to come down, but we have to have the
20 ability to go back in time and replicate our models.
21 So, lastly, there's some future policy
22 regarding credit scoring that we wanted to make sure
23 you were aware of. Today, unlike what people think,
24 your telephone bill, utility payments are not reported
25 to the credit bureau. Those are very powerful

1 predictors just like other payment methods of future
2 payment behavior and so of creditworthiness. And
3 there's been several studies that show that today.
4 And so with that, I would like to thank you
5 for giving me this opportunity and hopefully this gave
6 you a quick glimpse into the status of AI and how it's
7 being applied in credit scoring. Thank you.
8 (Applause.)
9 MS. GOLDMAN: And thank you, Angela, for
10 that very interesting presentation on credit scoring
11 and bringing in the related legal and policy issues.
12 So, next, Melissa McSherry will begin her
13 presentation.
14 MS. MCSHERRY: Thank you very much, and
15 thank you so much for having me today. I work with
16 Visa. Visa is the world's largest payment network,
17 and what that means is basically when you use a Visa
18 card your -- the merchant where you use the Visa card
19 basically calls their bank and says can I authorize
20 this transaction. And then Visa connects the
21 merchant's bank with your bank, who says yes or no,
22 that's a good transaction to authorize. And then that
23 signal goes back to the merchant, and all of that
24 happens if everything goes according to plan. All of
25 that happens almost instantaneously.

1 In that -- in that context, Visa is very --
2 we work very, very hard to make sure that the
3 transactions that are going through are legitimate
4 transactions or not fraudulent transactions. I think
5 fraud worldwide today is something like $600 billion,
6 so it's a lot of money, and we want to make sure that
7 we do as much as we can to help banks prevent any of
8 those fraudulent transactions from going through while
9 still making sure that all of the good transactions go
10 through. Basically, when you are actually the one
11 using your card, if you try to use it, that it
12 actually works.
13 So what I'm going to talk about today is one
14 way in which Visa is using AI, specifically computer
15 vision, to help us do that work of looking after and
16 preventing fraud on the Visa system.
17 So you might be asking what do puppy dogs
18 and blueberry muffins have to do with preventing
19 fraud. And I put this up just to sort of illustrate
20 both the challenges and the opportunity in computer
21 vision. So all of you could look at these pictures
22 and very easily discern what's a blueberry muffin and
23 what's a puppy dog. But using the techniques that
24 were available up until, you know, call it 2012, 2013,
25 this was actually a pretty hard problem for most

1 computers to solve. They would get it right about 75
2 percent of the time.
3 And in I think it was 2013 -- there’s a
4 competition that is run every year. And new
5 techniques, specifically things called convolutional
6 neural networks, started coming into play and started
7 dramatically improving the ability of computers to
8 correctly differentiate the muffin from the dog. And
9 so we're now at the point where these techniques can
10 generally differentiate not just muffins and dogs but
11 can differentiate different images about 97 percent of
12 time as opposed to 75 percent of the time, which is
13 really quite good.
14 If you think about human beings -- although
15 if you’re sitting there concentrating, you know, you
16 would always be accurate since most people don't
17 concentrate all the time and they do sometimes make
18 careless errors, human beings run at about 95 percent
19 of the time, right, when you give them a lot of
20 images. So this ability to look at a picture and so
21 to say this picture looks like this one, and this
22 other picture looks like this other one, this is one
23 of the applications of AI that has dramatically
24 improved.
25 And so now I’m going to talk a little bit

1 about how we use that application of that computer
2 vision application of AI in the context of fraud. So
3 just a couple of terms before we get started with this
4 particular example. First of all, what is a fraud
5 score? Like I said, whenever you use a card, Visa
6 basically attaches a score to the transaction that
7 goes to your bank that says how likely is it that we
8 think that this is actually you using your card versus
9 someone who’s trying to commit fraud using your card.
10 We provide that information to the bank so the bank
11 can make a decision about whether or not they want to
12 authorize the transaction.
13 And as you can imagine, we process a lot of
14 transactions, right? So that first thing we do in
15 every transaction is we attach a score from zero to
16 99. But then if we look across all of the
17 transactions, we can actually say, for instance, all
18 of the transactions in an hour, how many of them were
19 at, like, the highest score, got a score of 99? How
20 many of them were at the lowest score, got a score of
21 zero. And it's helpful to us to look at the
22 percentage of scores that are in each of those bands.
23 And the reason why is if you -- if we're
24 running along and 1 percent of the population is
25 getting the highest score, that 99, and it's nice and

1 steady and then all of a sudden like 10 percent of the
2 population is getting a 99, that means that probably
3 one of two things is happening. Either there's a
4 giant fraud attack, and there are fraudsters that are
5 trying to, in a very coordinated way, steal a lot
6 money, and this does happen sometimes, right, in which
7 case we need to intervene. And we typically intervene
8 by calling the banks that this is happening to.
9 Or there is something wrong with our models
10 or system or how we're processing things. And, again,
11 that's a situation in which we need to intervene and
12 we need to make sure that everything is actually
13 working as we expect. So not only do we look at the
14 fraud scores, we also look at the distribution of
15 those scores.
16 And so the next page, this is just -- this
17 is a made-up example, but I think it sort of
18 illustrates what's going on. So you can imagine that
19 this is a graph looking at the percentage of
20 transactions in a particular score band. And in this
21 particular case, I just did it over days, and it goes
22 up and down, and it goes up and down because, for
23 instance, the kinds of -- the mix of transactions that
24 you see on like a Friday night can be pretty different
25 than the mix of transactions you see on a Sunday

1 morning. And so the mix of transactions in a
2 particular score band can go up and down.
3 Now, if you look at this, it's pretty easy,
4 again like the puppy dogs and the muffins. It's
5 pretty easy to see that at the end there’s something
6 that looks a little bit different, right, that doesn't
7 -- that pattern doesn't look like all of the other
8 patterns that came before it.
9 And this is, again, pretty easy for everyone
10 in the audience to see that that’s different, but it's
11 actually kind of hard for the tools that we had prior
12 to those computer vision tools to pick this up, like
13 you can’t -- like a traditional control chart, it’s
14 hard to write a rule that will get this because the
15 actual numbers are sort of -- they’re inside the range
16 of the historical range, they’re going up, they’re
17 going down. They’re not -- it's just -- it’s hard to
18 write the rules. But, again, it's easy to see it
19 using computer vision tools.
20 And so what the computer vision tools let us
21 do is basically do what a person would do in terms of
22 looking at this and seeing a pattern that's different.
23 But the computer vision tools let us do that every
24 hour of every day. I mean, the computer doesn't get
25 tired and people do, like, they need to go do

1 something else other than look at charts all day.
2 It lets us look at hundreds of metrics, not
3 just one, right? And if you think about this, this is
4 a pretty simple chart that I put up here, right? This
5 is basically one-dimensional, right? We sort of look
6 at the scores, versus one-dimension. And so it's easy
7 to see the variation. If I had put a chart up here
8 that had multiple dimensions, like we were varying a
9 couple things at the same time, that very quickly gets
10 really hard, even for people, to see the differences.
11 But, again, the computer vision techniques that we've
12 been talking about can pick those variations up pretty
13 quickly and can identify those. So we can not only
14 monitor what's going on versus one dimension, we can
15 monitor what's going on versus multiple dimensions,
16 and it makes our monitoring that much better and that
17 much faster.
18 So just a quick explanation of how we've
19 applied this in our particular situation. Basically,
20 we built a model that looks at the distribution of
21 each of those score bands that we just talked about,
22 so, you know, for instance, scores of 10 to 19, right,
23 so it does this for each score band. And it looks at
24 those distributions for a five-hour period over each
25 of the last 120 days. Right, so this is lots of data

1 that’s coming in. Think of the computer as looking at
2 a chart, an hourly chart over the 120 days.
3 And from that, it forms an expectation of
4 what the current five-hour period is going to be,
5 right? Is the score -- is the distribution going to
6 be going up and then down? Is it going to be going
7 down -- you know, down and then up? Is it going to be
8 going, you know, one direction -- it forms an
9 expectation. And, then, and this is the part that
10 relates back to the puppy dogs and the muffins, it
11 looks at the actual picture and it compares it to its
12 expectation that it created based on the last 120
13 days, right?
14 And so on the top row, we see on the right
15 is sort of what we would expect, right, for this time
16 period from the data that’s come in over the last 120
17 days. And what we see on the left is what actually
18 came in. In those two pictures, the computer would
19 say, yep, those two things -- they look similar,
20 they're both blueberry muffins or they’re both puppy
21 dogs, right?
22 But on the lower band, what we see is the
23 expectation for the particular time period that we're
24 looking at is just that the scores will be going up
25 during the time period. But what we actually see is

1 that they're going up and then coming back down. And
2 the computer at that point says, no, no, no, these do
3 not look like they're the same. This is not --
4 something is not matching here.
5 And that, then, causes the system to
6 generate an alert and say, hey, a person, a human
7 being, needs to go look at this, right? It might be
8 that it's fine. It might be that it's just, I don't
9 know, Black Friday, right, and so all kinds of things
10 are different. Or it might be that there is an actual
11 problem and we need to get engaged and figure out what
12 the problem is, and we need to figure that out
13 promptly.
14 So in this particular case, what's going on
15 is the computer is basically taking a lot of work that
16 might have been kind of boring and tedious for the
17 people and doing the boring and tedious part and then
18 just pulling out the things that are interesting and
19 require human intervention so that human can then go
20 and figure out what we actually need to do
21 differently.
22 One thing I just want to call out about this
23 particular example is, you know, so Visa is using a
24 lot of different AI techniques across a lot of
25 different places in our system. These particular

1 techniques are probably a little bit more, you know, a
2 little bit more further along and more developed than
3 some of the most cutting-edge techniques, but they're
4 still -- you know, they're still on the front end of
5 being applied and serve real production environments.
6 And one of the reasons that we started with
7 something like a monitoring example, right, where
8 we're trying to monitor our own performance as opposed
9 to exposing this to consumers, was sometimes when we
10 implement new techniques in a production environment,
11 sort of outside of a laboratory, things don't work
12 exactly the way you expected them to.
13 And so we wanted to, in this particular
14 case, get a fair amount of experience working with
15 this, some of these cutting-edge techniques, in an
16 environment that was -- that where if they didn't work
17 exactly the way we expected them to, you know, the
18 impact would just be on us, like we would identify a
19 bunch of things we needed to look at that maybe we
20 didn't need to look at as opposed to the impact would
21 be on consumers.
22 And so, you know, as we talk about these
23 techniques, I think there is enormous promise. You
24 know, I consistently find that models used -- models
25 built using many of these techniques consistently

1 outperform other types of models. But I think it's
2 also important that we develop the practical skills
3 and how do we apply them, how do we understand them,
4 how do we interpret them, how do we make sure that
5 they're doing exactly what we think they're doing as
6 we go forward.
7 So with that, thank you guys very much. I
8 really appreciate it.
9 (Applause.)
10 MS. GOLDMAN: Thank you for that very
11 interesting presentation on how Visa is monitoring for
12 fraud.
13 Okay. And next we're going to go into some
14 medical uses of artificial intelligence, and we'll
15 begin with Dr. Michael Abramoff, who will look at
16 recent developments in that area.
17 DR. ABRAMOFF: Anyway, thanks so much for
18 inviting me, having me over. I'm both -- I have a
19 long history in computer science and AI, and it seems
20 that some others also have mentioned that they have
21 been doing this a while. And you can sort of see my
22 age from the fact that my master's thesis in 1989 was
23 on neural networks to simulate the brain. And so I've
24 been working on this for a while.
25 I'm also a professor of engineering and also

1 of ophthalmology and I'm a practicing clinician, as
2 well as the founder and CEO of IDx, which is the
3 company that had the first autonomous AI approved by
4 the FDA recently, so it's actually being used on
5 patients.
6 And so I want to talk a little bit about the
7 background of why AI in healthcare and specifically in
8 diabetes and specifically in diabetic retinopathy.
9 This is the most important cause of blindness, the
10 most important complication for people with diabetes,
11 that's what they most fear more than death or
12 amputation, they fear blindness.
13 And so we know very well what to do about
14 diabetic retinopathy, this complication. I mean, when
15 I see my patients, I know how to treat them, how to
16 operate them, how to manage them. The problem is
17 primarily that we don't find these patients. And so a
18 so-called diabetic eye exam is performed maybe 20 to
19 30 percent of cases because people don't have
20 symptoms, and so we need to look at the retina,
21 clinicians like me, and that doesn't happen.
22 It's mostly because it's really hard to get
23 an appointment with me, which is necessary for this to
24 happen. So the idea is, hey, let's use AI and imaging
25 to make this better. So this is how it works. I'm

1 not showing a demo, even though it would be only a
2 minute or two, because this is not the appropriate
3 context for that. But it's an autonomous diagnostic
4 AI system. It means it gets a point-of-care result in
5 minutes, but more importantly, there's no human
6 reviewer oversight, so no doctor ever looks at the
7 clinical result. The clinical diagnosis is being made
8 without a physician.
9 It means that you can now shift specialty
10 diagnostics like what I do as a specialist in an
11 academic hospital to primary care and retail clinics,
12 which, of course, increases the ease for which
13 patients can undergo this exam, and you can also do
14 something about cost of healthcare. Thank you.
15 It requires, right, a robotic camera because
16 you want to make sure you can do this exam on the vast
17 majority of patients, not just a few. It needs
18 assistive AI for the operator. We will not go into
19 that. And what it requires is a high school
20 graduation for that operator. And it's very important
21 that you need clinical proof that it's safe for
22 patients, right, and we'll go into that in more
23 detail.
24 And so like I said, I've been doing this for
25 a while and, you know, early on I said, hey, here's an

1 algorithm, in 2000, it can do it, let's just bring
2 this into practice, and that's, of course, not how it
3 works. You need to do a lot of science, and then you
4 also need to convince the FDA that this is safe, as
5 well as patients and physicians. And I don't show it
6 on the slide, but my nickname is actually The
7 Retinator, like a terminator, because in 2010 my
8 colleagues were thinking, hey, he's like a terminator,
9 he will destroy jobs, and he's also not being safe for
10 patients. And now they think very differently, but it
11 can show you how this fear of AI, you know, is not
12 new. And it's there and it's real, and so we also
13 need to manage that.
14 But, anyway, back to what happened is you do
15 science, and then for many years, you do more science
16 and more science, and you get grants -- thank you --
17 and NSF grants -- thank you, and many other grants,
18 and then more study sections, but eventually you get
19 to a point where -- we knew that the open source
20 wasn't going to work, so you need to go through the
21 FDA, raise money to go through the FDA because it took
22 us $22 million to do this, and then eventually you
23 build a company to do all of that.
24 And so one of the things that happened
25 during that path was that traditionally we use certain

1 features for essentially what we now call AI, and I
2 like the wave of AI so I'm calling it that, but we
3 took a sort of different approach because given the
4 experience in neuroscience, we tried to mimic the
5 brain of clinicians and say, well, clinicians do it
6 this way, why don't we build a computer that does it
7 the same way.
8 And there's a number of advantages that we
9 now realize that were sort of not even thought about
10 when we did it. And so we built detectors for each of
11 the different visions that you can see in the image of
12 someone with diabetic retinopathy. And on the right,
13 I show this sort of process where the orange images
14 are retinal images, and then you can detect different
15 diseases.
16 It's like the puppy images and the cookie
17 images that were just shown. We build detectors for
18 the eyes and for the raisins and other aspects of
19 that. And by now, it's being used in clinic.
20 Actually, patients are being diagnosed by the
21 clinicians, but again, no physician oversight.
22 So there was a scientific stage, I already
23 talked about that, and we learned a lot from that,
24 like the insights from neuroscience and the evolution
25 of mammalian vision story. I cannot read the slides

1 over there, so I have to do it from the big screen.
2 There were insights from clinical evidence, and it's
3 really important.
4 You need to put your AI in a work flow and a
5 clinical work flow, the care pathway, and it needs to
6 fit there, fit with the preferred practice patterns,
7 but the evidence about certain treatments that we
8 already have, and also you need to start thinking
9 about how you actually validate an AI when typically
10 you compare it to humans, but we already know that
11 humans, clinicians like me and my colleagues, have a
12 sensitivity, meaning the ability to detect disease of
13 about 40 percent, so it's pretty low. So we're not
14 really very good at making the difference between
15 subtle degrees of diabetic retinopathy, of this
16 disease.
17 And so how do you compare an AI to imperfect
18 clinicians, imperfect truth? And it was a big
19 challenge that we needed to solve. And they have
20 insights from interpretation and then poorness of
21 image quality, which is easy to reach in a laboratory
22 setting but very hard to reach in a retail clinic like
23 Walgreens or CVS where there's no one with any retinal
24 imaging training.
25 Anyway, I already talked about this approach

1 to essentially basing it on how the visual cortex, the
2 brain of clinicians, solve this problem, and we
3 started to implement that. And that has now a sort of
4 number of advantages that we had to realize at the
5 time but are now pretty logical.
6 And so -- but before I explain it this way,
7 I want to say that we already did a clinical trial in
8 2014, where we showed that we did better than
9 clinicians. And we thought, well, that's important.
10 We do better than clinician, that should be enough.
11 And the FDA and we and I agree with them now, they
12 rejected this clinical trial, saying, well, this is
13 not good enough. You need to show it in the actual
14 environment where you want to use it.
15 So what we did for this clinical trial, it
16 was used in academic ophthalmology clinics where
17 there's experienced photographers, the patient
18 selection is a little bit different, and we showed
19 this result. They said you need to show it in primary
20 care, with the people who already work there, the
21 staff that's already there, which is typically high
22 school graduates and no formal training in any type of
23 retina or retinal imaging.
24 You need to also decide on the truth, and
25 clinicians are simply not good enough, so how do you

1 compare it, what do you compare it to, and the answer
2 to that was reading centers where it's very
3 standardized for over 40 years. And you need to do it
4 like I already said with the patients that are already
5 there in those primary care or retail settings, not
6 with a more selective subset of patients.
7 So that was a clear lesson, and so these are
8 the lessons we and also the FDA, I think, learned from
9 this authorization that we got in April of this year,
10 a lot of things, system validation, all sorts of rules
11 about that. You need the highest level truth so you
12 can compare clinicians and the AI and also say that AI
13 meets certain standards in terms of safety and
14 efficacy.
15 And also I already talked about the system
16 as a whole. You do not evaluate it just as an AI and
17 reading images; it's actually a system, it's a robotic
18 camera with the operator, with the patients that are
19 already in primary care. And then you need to
20 preregister a trial, meaning you state what you're
21 going to analyze, what your hypothesis is, and you try
22 to prove or disprove that hypothesis about safety,
23 efficacy, and what the FDA and we thought was really
24 important, that the vast majority of patients need to
25 be able to undergo a diagnostic result. It's

1 relatively easy to make an AI that does really well on
2 a subset of about 10 percent of patients, but that's
3 not enough. You need to do it on the vast majority of
4 patients.
5 I will not talk about this slide. I put
6 these slides together two weeks ago. When I saw the
7 other slides, I realized this is not really the
8 subject of this meeting. This is more regulatory
9 stuff. But, anyway, so it cleared the path for
10 autonomous AI in general. So it took us a long time
11 to do this but now essentially the rules are set for
12 how you prove autonomous AI making these autonomous
13 decisions. And here are some of the implications
14 already talked about, explainability is now really
15 important.
16 And there's a number of advantages that were
17 already discussed, but we actually show that in
18 scientific studies and other groups have now confirmed
19 it. AI avoids racial and ethnic bias because by doing
20 a design this way, we explain it's based on detectors,
21 it's based on lesions that we already know about for
22 150 years, clinicians have been using.
23 When I look at a patient, I look for
24 hemorrhages, for example, and I don't care whether
25 that patient is from Iceland or Kenya, it doesn't

1 matter. If they have the hemorrhage, they have the
2 disease, and the AI does that the same way. But you
3 also avoid the lack of robustness that leads to
4 catastrophic failure. We talked about adversarial
5 images earlier. Well, we look at it as very small
6 perturbations in the image that are not visible to
7 humans that are not visible to an explainable AI, but
8 that CNNs -- typical use of CNNs are very vulnerable
9 to, and we show that essentially you have catastrophic
10 failure in 90 percent or more of cases.
11 I have two minutes left, right? And like
12 was said already, preregistered clinical trial is
13 really important to prove the safety. It's
14 essentially how we approve drugs, as far as the trial
15 is concerned. And then it needs to fit into the
16 clinic. We already talked about that.
17 And so I will move to the next slide, which
18 is, well, what are the implications for others
19 following us, and I think it’s very important. It
20 took us a lot of time, but it doesn't mean that others
21 will have the same problem. I think the rules are set
22 now. On the right, you see some implications of doing
23 it the wrong way. I mean, Bad Blood, you probably saw
24 the book, and that’s not how we want to do
25 improvements in healthcare and use technology in

1 healthcare.
2 And one of our competitors had said the
3 following, you know, it doesn't matter if you harm
4 some patients or harm something along the way to
5 improving technology and using technology in, for
6 example, healthcare and this autonomous driving. This
7 appeared in the New York a few weeks ago. So it’s
8 very -- it’s very cogent right now to do this in the
9 right and safe way. So we need to agree on
10 definitions and nomenclature.
11 You know, technology used in a lab does not
12 directly transfer to what we do in healthcare, and
13 it's very important. Patient safety is very
14 paramount. And if we don't do it right, there will be
15 pushback and we’ll lose all the advantages that AI can
16 have in healthcare for better quality, for better --
17 you know, lower costs, and for better accessibility,
18 meaning easier for patients to have it.
19 So, again, I think these are the lessons we
20 learned, that the FDA learned, and I think it will be
21 very important going forward that if you do autonomous
22 AI, we follow these lessons. Thank you.
23 (Applause.)
24 MS. GOLDMAN: And thank you, Dr. Abramoff,
25 for that very interesting discussion of how you

1 developed autonomous AI and got FDA approval for your
2 system. Thank you so much.
3 And now we're going to have Teresa Zayas
4 Caban, who will continue to look at the use of AI in
5 the medical field.
6 MS. CABAN: Hi. Good morning, everyone.
7 Very happy to be here and join you to discuss
8 opportunities and considerations of the use of AI in
9 health and healthcare and briefly discuss some
10 activities that my office has engaged in as well as
11 some of our sister agencies in the U.S. Department of
12 Health and Human Services.
13 A little bit of background before I get
14 started. I work at the Office of the National
15 Coordinator for Health Information Technology.
16 That's a staff division within the Office of the
17 Secretary of the U.S. Department of Health and Human
18 Services. Our charge has been really to facilitate
19 the implementation and adoption of electronic health
20 record systems.
21 OMC was created by executive order under the
22 Bush Administration and statutorily authorized with
23 the passage of the Recovery Act. There's a big
24 section in the Recovery Act called the HITECH Act,
25 which created a bunch of different things. One of

1 them you may have heard of. It created an incentive
2 program for eligible hospitals and providers to adopt
3 and meaningfully use an electronic health record
4 system. It also created a certification program which
5 the office I work in runs to certify -- to ensure that
6 an electronic health record system includes certain
7 functionality.
8 So with that backdrop, the number of
9 electronic health record systems across the U.S. has
10 increased significantly, with about 90-some-odd
11 percent adoption in hospitals and close to that in
12 ambulatory practices. And in 2016 -- in December
13 2016, the 21st Century Cures Act was passed, which
14 sort of shifted our direction a little bit to focus on
15 now we have these systems in place, how do we make
16 them talk to each other.
17 So our priorities since then have been to
18 focus on interoperability of electronic health record
19 systems and health IT systems, facilitating the
20 liquidity of health data to enable effective and
21 efficient healthcare delivery as well as reducing
22 provider burden or improving usability of these
23 systems so clinicians have an easier time using them
24 in practice.
25 So how do we get into AI? Today, I'm going

1 to talk specifically about a report that we released
2 in collaboration with the Agency for Healthcare
3 Research and Quality and the Robert Wood Johnson
4 Foundation that was conducted by an advisory group
5 called JASON. And I'll walk you through the goals of
6 the report and some of the recommendations that came
7 from it.
8 Leading up to the study, as you may have
9 heard earlier in this panel and earlier today, there's
10 been a lot of progress in AI broadly with the increase
11 in compute and the increase in large data sets that
12 are high quality and well-labeled, a lot of strides
13 have been made in machine learning and artificial
14 intelligence. So with that, we saw also an increase
15 in clinical applications.
16 And so one of them you may have heard about
17 is in dermatology. And it looks like -- and the most
18 recent one -- my slides are a little changed -- the
19 most recent one is an application developed by Google,
20 really looking at whether an AI application can detect
21 metastatic cancer from a cancer that has not spread.
22 And they've been able to demonstrate this successfully
23 99 percent of the time. This tool that they've
24 developed has actually detected metastatic cancer and
25 distinguished it from a slide that doesn't have

1 cancer.
2 It was also able to accurately pinpoint the
3 locations of both cancers and observe lesions that,
4 frankly, a pathologist would just not be able to
5 detect with the naked eye. These tools really have
6 the potential to improve care but may require
7 adaptation for successful clinical use. And it is
8 important for them to be deemed effective and be
9 spread across healthcare and different applications,
10 that the technical soundness of their algorithms be
11 tested and demonstrated, that they perform at least
12 well as the current standard of clinical care. They
13 need to be tested across a wide range of situations
14 and really need to provide improvement, whether that
15 be in patient outcomes, practicality of use, or
16 reduced cost.
17 I was at the American Medical Informatics
18 Association's annual symposium last week where Jess
19 Mega from Verily Life Sciences gave the opening
20 keynote remarks, and she talked specifically about the
21 need for rigorous testing and appropriate development
22 and application of AI tools for them to be successful
23 and broadly adopted and used in health and healthcare.
24 Before I go over the goals of the report, I
25 wanted to briefly mention that this is not our first

1 collaboration with JASON. So the Agency for
2 Healthcare Research and Quality and Robert Wood
3 Johnson have previously collaborated on two studies
4 with this group. JASON is an independent group of
5 scientists that have been advising the Executive
6 Branch of the Federal Government for many years. And
7 we specifically engaged them in a study entitled "A
8 Robust Health Data Infrastructure," which helped
9 inform some of our office's direction in terms of
10 interoperability a few years ago.
11 We also engaged them in a separate study
12 called "Data for Individual Health," which looked at
13 how EHRs and health IT could support individual
14 health, allowing individuals to have access to their
15 own health data. And this has actually -- the
16 recommendations from this report have helped spur the
17 health ecosystem we currently have. A notable example
18 is Apple's use of ONC-recognized standards to
19 implement their health app, which has now enabled
20 individuals to download health data to their iPhones
21 from a whole host of healthcare provider systems.
22 This third collaboration is the focus of
23 this presentation and began a little over a year ago
24 when we asked JASON to consider how AI could help
25 shape the future of public health, community health,

1 and healthcare delivery. The report focuses on the
2 technical capabilities, limitations, and applications
3 that can be realized in the next ten years.
4 We asked JASON to consider the
5 opportunities, considerations, and implementation
6 issues around the use of AI in health and healthcare.
7 So under opportunities, there were things -- questions
8 that they asked or looked at where ways where AI may
9 advance the improvement of health and healthcare,
10 evidence that currently exists regarding AI's
11 relevance for health and healthcare, most high value
12 areas, and what kinds of benefits can be defined and
13 assessed.
14 In terms of considerations, there were three
15 categories that we asked JASON to look at. One was
16 technical considerations; the other one ethical and
17 legal issues; and the last one, workforce issues,
18 which are very important if we're actually going to
19 see increased development of these applications and
20 their implementation across healthcare.
21 And in implementation, we really asked them
22 to look at other fields and what lessons could be
23 learned that would be relevant to the development and
24 implementation of AI in health and healthcare.
25 So what did they find? Essentially, JASON

1 concluded that the time may be right for the use of AI
2 in health for three reasons that are noted on this
3 slide. Namely, there’s frustration with the existing
4 medical systems, the ubiquity of smart devices, and
5 comfort with at-home services. JASON outlines a
6 series of findings and challenges and makes some
7 recommendations about how to successfully apply AI in
8 health and healthcare.
9 And I'll go over those quickly, and I have
10 included the link to the report so you can sort of
11 peruse that at your leisure, and I'm happy to answer
12 questions after the session today. So JASON found
13 that overall, AI's beginning to play a growing role in
14 transformative change now underway both in health and
15 healthcare, meaning in and outside of the clinical
16 setting.
17 So the first challenge they identified was
18 regarding acceptance of AI applications. And so they
19 really recommend supporting work to prepare AI results
20 for rigorous approval procedures, as well as creating
21 testing and validation approaches under conditions
22 that differ from those used for the training set.
23 With regards to leveraging personal network
24 devices, JASON recommends supporting development of AI
25 applications that can enhance performance of new

1 mobile monitoring devices and apps, developing the
2 necessary data infrastructure to capture the data
3 generated from smart devices to support AI
4 applications and requiring development approaches to
5 ensure privacy and transparency of data use, which is
6 a little bit of what Dr. Kearns alluded to in his
7 remarks earlier this morning.
8 With regards to the issues around training
9 data sets, they really recommend the development of
10 research, with development and access to research data
11 of labeled and unlabeled health data to support
12 development of AI applications. They suggest that new
13 models are needed to incent the sharing of health data
14 and new paradigms of data ownership.
15 Some of you may have heard of a movement
16 called Open Science. So there's really an interest in
17 sharing research data sets, but then in healthcare
18 more specifically, there's privacy and security
19 considerations attached to the data. So we need to
20 rethink under what circumstances we can share data to
21 enable both discovery, as well as development of these
22 applications, and validation of these applications so
23 they can be more broadly used.
24 They also made some recommendations
25 regarding collecting data that are relevant to health

1 but are not systematically collected or integrated
2 into clinical care. So one example is environmental
3 exposure data. But, today, your health is determined
4 mostly by where you live more so than your genome. So
5 we really need to think about what kinds of data are
6 important to health and health care and how we make
7 use of those data and include them into machine
8 learning and AI applications so we make the right
9 kinds of predictions to support whether it be
10 prevention, diagnosis, or treatment.
11 They really emphasized building on the
12 successes of other domains through competitions, for
13 example, as well as understanding the limitations of
14 AI methods and how they can be applied. They talked
15 about guarding against proliferation of misinformation
16 in this emerging field. As you can imagine, there's a
17 lot of hype about AI generally and specifically in
18 health and health data. So wading through that and
19 ensuring transparency, as well as endorsing best
20 practices by learned bodies.
21 So since I'm short on time, suffice to say
22 there's a lot of possibilities, there’s emerging
23 applications in health and healthcare, and they range
24 from public health to clinical health, as well as
25 prevention and treatment. Our role is really to work

1 with other agencies to identify what those
2 possibilities are. Our focus is on making data
3 interoperable, to be able to support a development of
4 AI and understanding the data infrastructure issues
5 and what kinds of standards are needed to enable this
6 vision.
7 And before I close up, I did want to mention
8 two efforts that I thought would be of interest to
9 this audience. So Gina Tourassi heads Health Science
10 Data Institute in the Oak Ridge National Lab that has
11 two big collaborations -- one with the National Cancer
12 Institute and another one with the Veterans Health
13 Administration -- that are really meant to leverage
14 both the compute power and the methodological
15 background that folks at Department of Energy have
16 with the data sources, as well as the research
17 questions and health questions that folks on the other
18 end have to enable new solutions.
19 With that, I'll stop.
20 MS. GOLDMAN: Thank you, Teresa. We
21 certainly appreciate your discussion of those issues
22 in the field of medicine.
23 (Applause.)
24 MS. GOLDMAN: And it’s a great place to
25 begin the discussion section now. So we've had a lot

1 of discussion of the use of AI in different
2 situations. But at this point, I'd like to put the
3 question squarely on the table. Under what
4 circumstances do our panelists think that it might be
5 better to use artificial intelligence technologies,
6 broadly speaking, rather than traditional algorithms
7 and vice versa? And in considering that, is the
8 selection of the technology generally based on
9 technical considerations or the purpose of the
10 analysis, or are there other practical policy or
11 ethical issues that might add to the decision, some of
12 which we've certainly heard about already today?
13 So if anybody would like to address that
14 question, please turn your table tent on the side.
15 So is there anyone -- okay, you would like
16 to? Go ahead, then. Thank you.
17 MR. RAO: So when we look at when we would
18 use AI versus traditional software programming
19 techniques, the easiest cases for us are anything that
20 -- you need a pattern for -- as he mentioned, we're
21 looking for pattern recognition, so the technical
22 subject matter of what we are trying to do has to be
23 something that we can -- is repeatable and we can
24 train for. So we have to be able to have data that
25 can reveal the problem over and over again so we can

1 train the AI on it. So that's the kind of problem
2 that we can solve with AI. So for us, it has to fit
3 in that category.
4 If it's a very intuitive decision or a one-
5 off decision or something that’s not going to be
6 repeated, it's not a candidate for us to use AI for,
7 and that's still a candidate for what we refer to is
8 human assistance. So when we think about how to
9 design our software programming, we're looking at what
10 parts can we pull away that are the AI parts and what
11 parts are the parts that are probably always going to
12 be left up to the individual to add their value.
13 MS. GOLDMAN: Thank you.
14 MR. KAUTZ: Yeah, so there's a lot of work
15 and interest in human-in-the-loop systems, and that’s
16 probably actually the major category of deployed
17 applications, where we’re not -- it's a person working
18 together with an AI system. I mentioned in my talk
19 examples where people on their own, they simply can't
20 handle combinatorics of the problem, so that's a good
21 opportunity for using an AI system together with a
22 person.
23 And I think a number of the people here
24 have talked about these issues of fairness and
25 transparency. There's also some, you know, deep

1 ethical issues. So there has been work, particularly
2 actually in Japan, on robotic friends for the elderly.
3 So these are not truly artificial intelligence
4 systems. They're simulated animals or simulated
5 people that people with diminished capacity might
6 actually come to regard as friends and have an
7 emotional bond to. And I think that could be an
8 example of something we could do but we just should
9 not go down that path.
10 MS. GOLDMAN: Thank you.
11 Angela?
12 MS. GRANGER: Yeah, just to add to, you
13 know, the explainability side is very -- very
14 important, but also the ability to actually implement.
15 If you think about a lot of the techniques that have
16 been talked about, and neural nets, you know, I'll
17 just pick on because it was mentioned a few times,
18 that's been around a long time. And in our industry
19 in particular, one of the reasons it hasn't -- it
20 never took off is because the implementation was more
21 difficult.
22 And so the technology today is there, so
23 when you're doing your research and your analysis, you
24 always have to think about the application and whether
25 or not it can actually be used in production.

1 MS. MCSHERRY: You know, just to build on
2 what some of the other speakers have said, we are
3 consistently finding when we look at AI techniques --
4 and I'll compare to that what I might think of as more
5 traditional techniques like logistic regression or
6 gradient boosted trees, but when we look at AI
7 techniques, we are consistently finding that those
8 models are outperforming the more traditional
9 techniques.
10 I think that the -- you know, one of the key
11 challenges is making sure that you have enough data so
12 that the models are not overfit. I think -- I don't
13 know that AI necessarily is inherently more likely to
14 be overfit, but because people are less experienced
15 using it, the human beings are more susceptible to
16 overfitting their models. There are good rules of
17 thumb for how to avoid overfit in something like
18 logistic regression, and the rules of thumb are maybe
19 not as well developed with AI techniques.
20 I'm optimistic, though, as more people start
21 building these models, those rules of thumb will come
22 as well. So I think, you know, having enough data is
23 one of the key considerations.
24 And then as Angela said, you need to have
25 enough, you know, computing power, right? So these

1 are computationally expensive models to build, and
2 depending on how you structure them, they can be
3 computationally expensive to run. And as long as you
4 have enough computing power, that's not an issue, but
5 one definitely does need to have enough power.
6 MS. GOLDMAN: Thank you. That’s very
7 helpful.
8 DR. ABRAMOFF: Yeah, it’s interesting, I
9 think where you need performance, especially in
10 autonomous AI, you need, you know, techniques that
11 work. And, so for instance, really the techniques
12 that work, and it seems to be that AI is now starting
13 to be essentially whole-vector-based deep learning
14 where you don't know what it's doing.
15 I don't think that’s what AI is. These deep
16 learning or convolutional neural networks are a
17 technique. There's many different machine-learning
18 techniques that you can all use, and what you saw --
19 what we do is we combine convolutional neural networks
20 as detectors and there’s sort of a hybrid rule-based
21 system over that in another AI to combine it into an
22 actual dichotomous output.
23 So there's many different ways, but you
24 still call the entire thing an AI. I think that's
25 valid. And so, for me, it's higher performance, the

1 better you understand it, the better, but AI doesn't
2 necessarily mean that you don't understand it. We
3 showed that we have AI that you can clearly understand
4 exactly what it does.
5 MS. CABAN: So quickly to build on others'
6 comments, I would say that in healthcare, it's not
7 like there's this set number of circumstances under
8 which AI should be used, but there's certainly some
9 parameters that should be kind of guiding principles
10 that I alluded to during my remarks and that Michael
11 was just alluding to.
12 You really need to be able to demonstrate
13 that this is as effective or more effective than
14 standard clinical practice. And it really needs to
15 lead to better outcomes. All right? And so if
16 there's enough testing and transparency around
17 whatever AI tool or application is being developed, so
18 long as it's better than the current standard of care
19 and it's been shown to improve something that really
20 needs to be -- that's right for automation.
21 I really see AI as a tool that can help
22 augment clinical care. Clinicians are extremely busy.
23 There's a lot of data, there's a lot of knowledge that
24 they need to wade through to provide effective care,
25 so think about how AI can help them do that in an

1 unobtrusive manner and in a way that reduces a burden
2 on them to be able to practice.
3 DR. KEELING: Thank you. So the next
4 question is how accurate are the algorithms in AI
5 tools that we've heard about this morning. And if
6 there is a wide range of accuracy, why is that so?
7 And, also, is the accuracy related to the nature of
8 the tool, the question being asked, or the data being
9 used?
10 MS. MCSHERRY: So, look, I think, again, in
11 our experience, the AI -- the models that we build
12 with AI, when the competence of the practitioner and
13 the data being made available is the same, and we
14 generally don't suffer from a shortage of data, just
15 given what we do, in those cases, we generally find
16 the AI models to be more accurate. But those two sort
17 of -- when these two things are the same, the data
18 involved and the competence of the practitioner, those
19 are often not actually the same in the real world.
20 And so I think that the algorithms
21 themselves are -- again, my experience -- very
22 powerful and very effective. And we -- but the models
23 that come out the other side can have a wide range of
24 accuracy because you may or may not have adequate data
25 that’s relevant to the problem being solved and you

1 may or may not have a person who's building the model
2 who is really effective at structuring that model to
3 get the best possible outcome.
4 So, you know, when we think about the
5 outputs of the models, there can be a wide range, but
6 my experience has been that has much more to do with
7 the data that's available and the sort of technical
8 competence of the person building the model than it
9 does the actual algorithms, which again, when we do
10 head-to-head tests seemed to pretty consistently
11 produce outcomes that are better using the advanced AI
12 techniques.
13 MS. GRANGER: Yeah, and just to add on to
14 that, there's -- you know, credit scoring has been
15 done for many, many years, so it's a very well
16 established predictive use of analytics. And so the
17 lift that you see isn't -- not probably as great as it
18 is in something that's more a greenfield that hasn't
19 been done for as long as credit scoring has been.
20 But when I mentioned earlier in our
21 particular study we saw a 5 percent lift in using some
22 of the more newer techniques outside of regression,
23 what I didn't mention is if you add new data in,
24 you'll also see another 5 percent lift in performance,
25 right? So the data becomes very valuable, regardless

1 of the methodology being used.
2 DR. ABRAMOFF: It's probably the most
3 challenging problem in medicine, in medical AI, is
4 that what do you compare it to. I and my colleagues
5 differ in about 30 percent of cases. And so if you
6 compare an AI to an individual clinician, when do you
7 know the AI is right and when do you know the
8 clinician is wrong? You will never say that.
9 And so averaging clinicians will not work
10 much better either. And so we look for ways of doing
11 better. And you can see from our actual trials that
12 we had really good performance -- 97 percent
13 sensitivity catching the disease -- on a data set that
14 was not ultimately to be used in a clinical trial that
15 the FDA authorizes on, where we shot 87 percent
16 sensitivity, the same system. So that risk can be
17 perceived to be very different depending on what you
18 compare it to. And I think it's really, really
19 important that you compare it to the best standard out
20 there, which is usually better than an individual
21 clinician or even a group of clinicians. But that's a
22 challenge that is not really resolved.
23 MS. GOLDMAN: Okay, so I would like to ask
24 an audience question at this point. I just want to
25 say that we're not going to get to all of the audience

1 questions, but we're not going to get to all of the
2 moderator's questions either. And we will hang onto
3 these questions and keep them in the FTC record.
4 But I'll start with this one. What, if any,
5 efforts do you make to improve your applications of AI
6 after implementation? Do you test for anomalies? Do
7 any third parties review your implementations to
8 provide oversight as you identify problems?
9 MS. CABAN: So I'll make a general comment,
10 not specific to AI, but like anything else, you have
11 to keep evaluating and testing, so it's part of this
12 continual life cycle, engineering life cycle, whatever
13 you call it in whatever field or discipline you're in.
14 So you have to do that with AI, same as you would with
15 any tool.
16 In healthcare in particular, after something
17 is implemented, you need to make sure it's working as
18 intended and not leading to unintended consequences,
19 undue harm, slower processes, or less effectiveness in
20 care.
21 DR. ABRAMOFF: Yeah, the FDA required us
22 to build a whole system for continuous efficacy
23 monitoring, meaning we have to consistently monitor
24 that it's up to what we did in the clinical trial.
25 MS. MCSHERRY: Yeah, I mean, just to pile

1 onto that, I think it's basic good practice that you
2 have to monitor a model. And that’s not -- again,
3 that's not specific to the technique, like you need to
4 do that with any model, whether it’s logistic
5 regression or gradient boosting tree or deep learning
6 or CNN or LSTM or really any algorithm. Like, if you
7 don't monitor the performance of the model, eventually
8 it will degrade and you won't catch it and then you’ll
9 make mistakes.
10 MS. GRANGER: Yeah, pretty much the same
11 thing I was going to say. Not only that, it's also
12 regulated for us to need to monitor the model and show
13 performance.
14 MR. RAO: I think in addition to the regular
15 engineering testing, I think for us the new part about
16 AI is understanding that we have to test for inherent
17 bias in the data set, so that was not something that
18 Adobe did traditionally in its software practices.
19 And we wrote an algorithm in PhotoShop that was not
20 something we had to think about, but now when we train
21 data to sort out pictures and answer queries and
22 understand content, we actually have an explicit
23 second step of understanding and testing for implicit
24 bias. So that's new because of AI.
25 MS. GOLDMAN: Thank you.

1 MR KEELING: So my question asked, what
2 factors have facilitated the development and
3 advancement of these technologies? Have certain
4 resources and policies facilitated their development?
5 MS. MCSHERRY: Yeah, look, I think that
6 there are a couple things out there that have been
7 very helpful. First, for us at least, the
8 availability of open source algorithms and the
9 availability of open source data sets has been super
10 helpful. I actually have a person on my team who is a
11 veteran of 20 years of using traditional techniques.
12 And she built her first TensorFlow model a couple
13 months ago, and I said, wow, that's great. And she
14 said, yeah, you can find anything on the internet
15 because, you know, she was able to find, you know,
16 basically everything she needed to go learn this new
17 advanced technique, because it's just all out there.
18 And so I think the availability, the robust
19 open source environment and the availability of open
20 source tools is something -- has certainly been
21 something that we have benefitted from greatly and
22 we're very supportive of.
23 MR. KAUTZ: There is also a big advance in
24 hardware around 2007 that made these techniques for
25 deep learning that date back to the '40s and then with

1 additional work done in the ‘80s suddenly scale to
2 real world problems. And this was the discovery by a
3 group of researchers that you could repurpose the
4 graphics processing units of computers that had been
5 developed for computer games and for computer graphics
6 and movies.
7 These were just the perfect things to use to
8 run neural nets. And they gave a 10,000-fold increase
9 in speed. And you very rarely get a five order
10 magnitude speed-up. And when that happens, suddenly
11 ideas that could only handle tiny problems, you know,
12 perhaps they could read a zip code, could scale
13 tremendously. So there is that kind of hardware
14 breakthrough.
15 More recently, companies -- Google,
16 Facebook, Intel, and ARM -- are all coming up with
17 further hardware advances that are tailored for
18 running deep learning systems. And nothing so far
19 will give a 10,000-fold speed-up that’s on the near-
20 term horizon, but perhaps with some radical new ideas
21 about analog circuits, we might see at some point to
22 the next decade another discontinuity in the
23 performance.
24 MR. RAO: Just on the legal side, what's
25 been helpful, especially for our neural nets, which

1 were trained on images and documents, is in the United
2 States we have fair use exception to the copyright
3 law, and we can use that to allow ourselves to and
4 other communities like us to access publicly available
5 works to train our machine learning.
6 In contrast, in Europe, they have a
7 copyright directive which currently prohibits that,
8 and it makes it much more difficult to get data to
9 train our neural networks from Europe, and there's
10 some momentum around changing that, but I do think
11 it's valuable to point out that the legislative
12 framework could also hinder or help development of ML
13 and neural networks.
14 DR. ABRAMOFF: Yeah, on the regulatory side,
15 I want to do a shout-out to the FDA because they have
16 been extremely understanding and willing to help and
17 make this happen, and now we have the first one
18 approved -- authorized, very careful -- this year. So
19 I think from the regulatory perspective, it's great.
20 I want to make another remark from the sort
21 of science funding perspective, I've been filing for
22 NSF and NIH. That's also really important starting
23 on, but more importantly, these algorithms existed
24 from Fukushima in the '80s. And I used deep learning,
25 you know, back propagation.

1 I think for us in healthcare, it's always
2 grappling with noisy, insufficient data and sensor
3 design in cameras, et cetera. It that’s what’s really
4 important because I think AI previously failed in
5 medicine, at least, because the inputs were actually
6 noisy. It was usually clinicians hearing patients
7 talk. We then typed it in, and that’s just not good
8 enough to have a really good performance. So the
9 problems we are now having with comparing to
10 clinicians are stemming from the fact that we’re so
11 good and that is because better sensory data is
12 available. A long story but...
13 MS. CABAN: Yeah, to add to Michael's
14 comment, in healthcare, we struggle with the data
15 quality, data completeness, and missing data. And so
16 that creates a unique set of considerations if these
17 applications or tools are going to be developed using
18 data that’s in electronic health record systems. And
19 there really is a need to better understand what it is
20 we can design with poor data quality and how far we
21 can stretch those models.
22 MS. GOLDMAN: Well, I really wish that we
23 could continue the discussion, but we are running out
24 of time now. So I would like to ask everyone to join
25 me in thanking our wonderful panel here.

1 (Applause.)
2 MS. GOLDMAN: And we'll now have a break for
3 lunch, and we'll be back after that at 1:15.
4 (End of Panel.)
5 (Lunch recess.)
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 PERSPECTIVES ON ETHICS AND COMMON PRINCIPLES
2 IN ALGORITHMS, ARTIFICIAL INTELLIGENCE, AND
3 PREDICTIVE ANALYSIS
4 MR. TRILLING: Welcome back from lunch. We
5 are about to resume the hearing. Our next panel will
6 discuss perspectives on ethics and common principles
7 in algorithms, artificial intelligence, and predictive
8 analytics. My name is Jim Trilling. I am an attorney
9 in the FTC's Division of Privacy and Identity
10 Protection, and I will be co-moderating the panel
11 along with Karen Goldman who, if you were tuned in
12 or attending this morning, you have already met.
13 Karen is an attorney in the FTC's Office of Policy
14 Planning.
15 We are pleased to have a great group of six
16 panelists to discuss ethics and common principles
17 related to artificial intelligence. The format for
18 this panel will be similar to the last one. Each
19 panelists will make a presentation and then we will
20 have a discussion about issues that are raised in the
21 presentations.
22 We again welcome questions from the
23 audience. Note cards are available for you to provide
24 questions if you want to write them down during the
25 panel.

1 I am briefly going to introduce our esteemed
2 panelists in the order in which they will be
3 presented. I am sorry, in the order in which they
4 will be presenting. You can find more detailed
5 information about each panelist in the biographies
6 that we have printed and made available on our
7 website.
8 Our first panelist is James Foulds, an
9 Assistant Professor at the University of Maryland,
10 Baltimore County. Following James will be Mark
11 MacCarthy, Senior Vice President for Public Policy at
12 the Software and Information Industry Association.
13 Then we will hear from Dr. Rumman Chowdhury, the
14 Global Lead for Responsible Artificial Intelligence at
15 Accenture; then Martin Wattenberg, a Senior Research
16 Scientists at Google; then Erika Brown Lee, a Senior
17 Vice President and Assistant General Counsel at
18 Mastercard; and, finally, from Naomi Lefkovitz, a
19 Senior Privacy Policy Advisor at the National
20 Institute of Standards and Technology.
21 With that, I will turn the microphone over
22 to Professor Foulds.
23 MR. FOULDS: It is great to be here. This
24 first presentation is on fairness and bias and machine
25 learning and artificial intelligence systems.

1 So let's make sure we are on the same page.
2 I want to briefly talk about what machine learning is.
3 So we are becoming increasingly aware that machine-
4 learning algorithms, which make predictions based on
5 data, are making a big impact on our lives. A common
6 example that we all deal with is credit scoring, so
7 predicting whether you will repay or default on a
8 loan.
9 So on the slide, we have a bit of an example
10 of how this might work. So you have some features for
11 the person. So for example, you would have the number
12 of late payments and the amount of credit used,
13 previous defaults, whether or not you are employed,
14 and so on. Then based on these features, you try to
15 make a prediction, in this case, whether you will
16 repay your loan or not. So the features, they are
17 called a feature vector or instance, and then the
18 thing you are trying to predict is called the class
19 label. So you try to predict the class label Y given
20 the features X.
21 So these models are trained using a bunch of
22 these feature vectors and they try to imitate what is
23 in the data set, and this is called classification.
24 This is an instance of supervised machine learning.
25 So it is supervised because the labels are provided.

1 So there is growing awareness that biases
2 inherent in these kinds of data sets can lead the
3 behavior of machine-learning algorithms to
4 discriminate against certain populations. There are a
5 number of high-profile papers and books on this
6 subject.
7 So for example, the Executive Office of the
8 previous administration published a report called Big
9 Data: A Report on Algorithmic Systems, Opportunity
10 and Civil Rights. And this was really a call to arms
11 to researchers in both computer science and law and
12 other disciplines to start thinking about these
13 problems. So they showed a number of, more or less,
14 hypothetical case studies about how things could go
15 wrong in terms of fairness and bias in machine
16 learning.
17 This book, Weapons of Mass Destruction, by
18 Cathy O'Neil, considers some of the same problem
19 domains, including housing and employment and credit
20 and criminal justice, and goes into greater detail on
21 a number of case studies.
22 One more book I want to point out is,
23 Algorithms of Oppression, by Cynthia Noble. So she
24 takes in intersectional feminist approach to
25 understanding this problem of bias and she looked

1 specifically at Google and how Google might, for
2 example, lead to problems of representation. So if
3 you search for the term “black women,” what kind of
4 results do you get compared to if you search for
5 “white women” or “white men.”
6 So there are also very serious real-world
7 applications where these problems are coming up.
8 There is the system that is already deployed today
9 called COMPAS, the Correctional Offender Management
10 Profiling for Alternative Sanctions. This system is
11 used to predict re-offending in the criminal justice
12 system, and it is being accused of being potentially
13 biased.
14 So there was an article by ProPublica
15 (Angwin, et al.) in 2016, and they found that this
16 COMPAS system tends to more frequently incorrectly
17 predict that black people will re-offend and end up
18 back in the criminal justice pipeline compared to
19 white people. And it found that the opposite happened
20 for white people, that you were more than twice as
21 likely to be incorrectly predicted that you would not
22 re-offend when you actually did if you were a white
23 person under this system.
24 So these findings are being disputed, at
25 least Northpointe would like to point out there are

1 other possible definitions of fairness that this
2 satisfies, but I do not think they dispute the main
3 claims that it does make these type of errors.
4 So let's look at an example to see how this
5 might actually happen. So I am going to show you an
6 example from a blog post by somebody called Rob Speer
7 and the blog post is called, "How to make a racist AI
8 without really trying." And so he is looking at an
9 application called sentiment analysis. So if you
10 think of reviews such as on Amazon or on Yelp where
11 there is a product or a service and you can type up
12 a review and post it online, we would like to predict
13 whether that review was positive, if you said that
14 was a good product or service or negative, if you
15 said that was a bad product or so. So that is a
16 sentiment label we would like to predict positive or
17 negative.
18 So once again you have feature vectors and
19 you would like to predict the class label. The
20 standard way to do this these days is to use something
21 called a word embedding, which automatically learns
22 for every word in the dictionary a feature vector.
23 And then given those feature vectors for the words, we
24 can try to predict the class label positive or
25 negative.

1 And so in this blog post, Rob Speer tried to
2 do this and he found that the system, just taking this
3 very standard approach, turned out to be horribly
4 biased. So you can look at the sentiment that the
5 model predicts for stereotypically black names and it
6 finds that the sentiment for those name is, on
7 average, substantially negative, whereas if you look
8 at the sentiment associated with stereotypically white
9 names, then the sentiment is extremely positive. And
10 the sentiment for Arab and Hispanic names is somewhere
11 in between. It is not as high as for white names.
12 Here is another example. So writers
13 recently reported that Amazon was trying to build an
14 internal tool for recruiting where they would like to
15 predict should we hire this person or not and they
16 found that this system was biased against women. So
17 it seems likely some of the same problems were the
18 cause of these issues that basically whatever was in
19 the data is somewhat discriminatory. For example, if
20 you tried to predict whether you will hire a person or
21 not and then you have mostly hired males in the past,
22 then the system is just going to encode that.
23 So where does this bias come from? So you
24 can look at this article by Barocas and Selbst, Big
25 Data's Disparate Impact. I will talk through some of

1 the reasons for bias that they point to. So for one,
2 data encodes societal prejudices. So we have already
3 seen an example of sentiment analysis where if you
4 just take data from the internet, let's say, and
5 people are just saying whatever they want to say, if
6 people are biased and they use that data, you are
7 going to encode those biases.
8 Data also encodes societal advantages and
9 disadvantages. Certain groups have performed poorly
10 in the past and the model is just going to learn that.
11 We also have, by definition, less data for
12 minorities. This could make you classify less
13 accurate for minority groups. And how you collect the
14 data, this can also be a problem. So if you imagine
15 we only collect data from smartphones, then you only
16 have data on people who have smartphones, so you are
17 going to ignore homeless people, for example, or
18 people who cannot afford a cell phone. This has
19 always been a problem in the past with polling.
20 If you do a phone poll, then you only find
21 people who have a phone in their home. In the early
22 days of polling that was a problem because it meant
23 that these were the wealthy people, you know, the
24 people who could afford a phone. But, nowadays, most
25 people do not even have a land line and so you are

1 getting a different demographic if you are calling
2 people who have land line phones.
3 You can also get cases of intentional
4 prejudice. This is sometimes called digital red-
5 lining. To hide that process, this is called masking.
6 There was a case of St. George's Hospital Medical
7 School -- this was in I think the late '70s, early
8 '80s when this happened. They encoded what they
9 believed was their own existing process for
10 determining whether they would accept a person into
11 their residency program and they made that system
12 specifically biased against women and minorities. The
13 people making those hiring decisions thought we should
14 not hire women because maybe they are going to get
15 pregnant or leave so we just will not hire them. So
16 they deliberately encoded that into their system.
17 And so it gets more complicated even if you
18 do not try to deliberately encode prejudice in your
19 system because every variable in your system, all of
20 your features in your feature vectors, are correlated
21 with your protective attributes like gender, race, and
22 age. It affects almost everything else about you. So
23 even if you leave those variables out, then you will,
24 by correlation, still learn some of those same
25 patterns.

1 So what do we do when we decide to model
2 fairness in an artificial intelligence context? So
3 this is very difficult to do. How do we nail down
4 what is fairness? You know, fairness is -- it is a
5 complicated sociotechnical, political, legal
6 construct, and nobody quite knows what it means. But
7 here are some considerations you might think about.
8 You might want to distinguish between the harms of
9 representation versus harms of outcome.
10 So when that sentiment analysis system -- a
11 harm of representation is where we see that the system
12 is biased against African Americans. And so in that
13 case, you may be offended by that. Maybe you were
14 upset that that is how you are being represented by
15 the system. But on the other hand, this may actually
16 affect an outcome that happens to you. So if I use
17 those same sentiment classifications or indeed the
18 features that drive them, then I may down weight your
19 CV if you are applying for a job.
20 Now, there are differences between equality
21 and fairness. So if we try to define fairness as
22 everything is equal for all groups, then we can run
23 into trouble if the groups are actually different.
24 You have to decide whether to model differences
25 between populations or not, should we treat these as

1 legitimate or should we encode them, and whether to
2 aim to correct biases in society as well as biases in
3 data. So you want to do something like affirmative
4 action.
5 So a related problem is explainability and
6 transparency. So many of these algorithms are
7 essentially inscrutable black boxes. So it is often
8 very hard to know what these methods are doing. So
9 sometimes there are legal reasons why you have to
10 provide some kind of explanation with these systems,
11 for example, credit scoring in the United States, and
12 then there is the GDPR protections in the European
13 Union.
14 The law does have some things to say about
15 it other than that. For example, we can just look to
16 Title VII and other anti-discrimination laws, which
17 prohibit employers and other parties from intentional
18 discrimination along lines of gender, race, national
19 origin, and religion.
20 The basic guidelines for this look at the
21 ratios of probabilities of a positive outcome like
22 hiring a person. And so if I hire all white people,
23 then if I hire black people at less than 80 percent of
24 that rate, then the law says that is an example of
25 discrimination.

1 The machine-learning community has also
2 tried to deal with these problems. So there has been
3 an explosion of research. It has been going on for at
4 least -- since 2012, but really it has received a lot
5 of attention since around 2016. They have been
6 cropping up new publication venues that are dedicated
7 to fairness and to related issues. There is the
8 FAT/ML workshop, Fairness, Accountability and
9 Transparency in ML, a spinoff ACM Conference, FAT*,
10 and there is a AAAI/ACM Conference on AI, Ethics and
11 Society that also has happened in the last two years.
12 In these research communities, there has been a lot of
13 work on defining fairness and algorithms that try to
14 enforce and to measure fairness.
15 Fairness can also be related to privacy,
16 which is another concern of the FTC. So for example,
17 if I have a system which assigns outcomes to people,
18 like a classifier, it may be possible, based on those
19 classifications, to determine which group you belong
20 to, are you a white male or -- and so on. And if that
21 is the case, then maybe even if our system was fair
22 then somebody could use that to discriminate later on.
23 For example, they could undo the fairness correction
24 that you have carefully down on your system. So this
25 is called the Untrusted Vendor Scenario (Dwork, et

1 al., 2012).
2 I would also like to point out that
3 fairness should be related to the study of fairness
4 in society, which has long been studied in literature
5 and feminism and especially intersectional feminism.
6 Intersectional feminism makes the argument that
7 systems of oppression built into society leads to
8 systemic disadvantages along intersection dimensions,
9 including gender, race, nationality, sexual
10 orientation, and so on.
11 So the argument is that if you are a
12 disabled Native American female, you are going to have
13 a very different experience than an able-bodied white
14 male. So, of course, that can be encoded in data and
15 that can lead to problems.
16 Now, there is a competing notion of fairness
17 called infra-marginality, which just argues that, you
18 know, different groups do have different distributions
19 over everything that happens to them, all of their
20 features and so perhaps we should define fairness not
21 as equality, but as the extent to which a system
22 biases above and beyond what is in society.
23 So in my research, I proposed a definition
24 of fairness which tries to look at both the privacy
25 aspect of fairness and intersectionality and it is

1 also related to fairness in the law, this 80 percent
2 rule where discrimination occurs with more than 80
3 percent difference between the groups.
4 So it has privacy and economic guarantees
5 and implements intersectionality and essentially it is
6 an extension of the 80 percent rule. But it allows a
7 sliding scale and it protects multiple protected
8 attributes and provides a privacy interpretation.
9 So that is it. Here are my contact details
10 if you would like to reach out to me. I have a
11 publically available pre-print of my work and another
12 pre-print is coming online soon. So thank you.
13 (Applause.)
14 MR. MACCARTHY: Hello. My name is Mark
15 MacCarthy. I am hoping that this clicker works.
16 So I am going to talk a little bit today
17 about some of the principles that my trade
18 association, SIIA, has put together. I want to start
19 off by saying we are not alone in this endeavor. The
20 Belmont principles, which many of you are familiar
21 with, the principles of respect for persons of
22 beneficence and justice were developed 30, 40 years
23 ago and they form the basis for the guidelines for
24 human experimentation and the IRB rules that many of
25 you are familiar with from an academic context.

1 The FAT/ML principles that were just
2 referred to are out there as well. ACM has a new code
3 of professional conduct for their members and for
4 software professionals. And our principles are in the
5 same ballpark. There are two others that I want to
6 mention, both of which have to do with human rights.
7 A group up at the Berkman Center at Harvard has put
8 together a series of very good applications of human
9 rights of some of these ethical principles and to hard
10 cases. And AccessNow has a similar document where
11 they talk about the importance of human rights in the
12 context of AI. So we are not alone in this endeavor.
13 Our principles are not original. You have
14 probably seen these concepts before. But before I get
15 into them, I want to say a word or two about when to
16 apply these principles because, after all, businesses
17 are engaged in lots of different practices and it may
18 not always be important to think about one from an
19 ethical point of view.
20 So the way I had sort of set it up is, when
21 the effect of business policy or procedure has large
22 effects on these values, these principles, then it is
23 important to pay enough attention to do an ethical
24 analysis and that is either positive or negative. If
25 there is a huge infringement of human rights, you have

1 to pay attention to that. If on the other hand your
2 policy or practice increases respect for human rights
3 and provides increased freedom of speech or increased
4 safety or further healthcare, then that is also
5 something that should be taken into consideration. It
6 is not just the negative stuff that you want to pay
7 attention to. So that is one.
8 The second point is that what is the status
9 of these principles, how should we think about them.
10 There is a continuum here from the kind of ACM
11 principles, which are really guides to individual
12 behavior, a code of professional responsibility. And
13 then that extends through guides to companies or self-
14 regulatory principles that might be enforced by a
15 group like the Digital Marketing Association and,
16 finally, soft law like the OECD principles that were
17 set up on fair information practices that eventually
18 became law in the European Union in 1985, and then
19 finally law itself.
20 I think we should think of these principles
21 as a guide for company action and not go farther down
22 the continuum. Part of the reason for that is most of
23 these principles are very, very abstract and the key
24 issues are really in the application of these
25 principles, not so much on the articulation of them.

1 And next steps really are not to further refine or
2 provide more detail on these principles. But it is to
3 apply them to particular cases. And that is where we
4 will find all the interesting ethical issues.
5 So for example, if you want to talk about
6 autonomous cars, the ethical issues involved are much
7 different from the ethical issues involved in
8 autonomous weapons. In the one case, you may need to
9 solve the trolley problem or at least assign
10 responsibility to people when something goes wrong.
11 In the other case, you may not even want to deploy
12 autonomous weapons unless you can figure out who is
13 responsible when a killer robot goes amuck.
14 So these are very, very different kinds of
15 ways of thinking about it. In other circumstances,
16 the companies disagree about how to apply these kind
17 of principles. So I do not think they are ready to go
18 beyond just guides for company action at this point.
19 So let’s get into it with that as the
20 background. Human rights. The idea is that when you
21 are engaged in various data practices, collecting
22 data, analyzing data, constructing models, you have to
23 respect internationally recognized principles of human
24 rights, and the sort of ethical thought behind that is
25 your behavior has to really respect the dignity and

1 autonomy of individuals. And you ought to not do that
2 in the abstract, but refer to the documents, the
3 guiding documents that have governed international law
4 for a couple of generations now.
5 And so which rights are we talking about?
6 Here is a sample from those international instruments,
7 the right to life, privacy, religion, property,
8 freedom of thought, and due process. I think
9 organizations should really be bound to validate those
10 internationally recognized aspects of human rights
11 law.
12 Justice. Here the real question is
13 distribution. When we start off with a principle that
14 individual people have a right to a fair share of the
15 benefits and burdens of social life and you want to
16 really be in a position where you are not engaged in
17 data practices that disproportionally disadvantage
18 vulnerable groups. In particular, you do not want
19 your data practices to result in applications that are
20 not available to all and are sort of intentionally or
21 even inadvertently restricted based on arbitrary and
22 irrelevant characteristics, which are race, ethnicity,
23 and gender or religion.
24 The organization should not be totally
25 indifferent to how their goods and services that are

1 produced are distributed. It should be a matter of
2 concern for them who benefits from their new
3 analytical services and products.
4 But that brings us to the important topic of
5 welfare. The whole goal of creating these new
6 processes and services is to increase human welfare,
7 and to the extent that you can do that through the
8 provision of public services or low cost and high-
9 quality goods and services, you have an ethical
10 obligation to do so.
11 The last grouping may be a little
12 unfamiliar. It is one of the standard ethical
13 theories. It is called virtue ethics. But the idea
14 is that you want your products and services to
15 contribute in some fashion to human flourishing. This
16 means that you are really trying to help people
17 individually and collectively to be the kind of people
18 who live well together in communities. And many of
19 these concepts are sort of old-fashioned. The words
20 that are used to describe this set of ethical
21 obligations are honesty, courage, modulation, self-
22 control, and the like.
23 But we all recognize that sometimes business
24 practices can discourage the development of those
25 virtues. All of the attention to things like the

1 addictive nature of some of the internet activities
2 leads you to think that maybe these devices are
3 teaching less in the way of honesty, courage,
4 moderation, and so on, and are more taking advantage
5 of people's weaknesses. So virtues are a very
6 important thing to pay attention to.
7 In many discussions, these four different
8 perspectives are thought of as sort of alternatives.
9 Pick one. Do you want to do justice or do you want to
10 do rights or welfare? Which is it? Our suggestion
11 was try to do them all. Treat them as a kind of
12 checklist and a set of guidelines to go through as you
13 are considering what needs to be done.
14 But the real issues here -- and this is to
15 repeat a point -- arise in specific domains. And I
16 think it is important to see how these principles are
17 applied in practice because that is where the key
18 ethical issues will really come to the fore.
19 So to talk about one that was raised before,
20 disparate impact analysis, as was mentioned, a key
21 part of assessing algorithms is to make sure that they
22 comply with the various statutory requirements
23 including the prohibitions on discrimination. There
24 are three stages of a disparate impact analysis. The
25 first is you have to take a look and see if your

1 algorithms are having a disproportionate adverse
2 impact on people. You have to see if there is a
3 legitimate purpose that is being served by this.
4 And then the third step is you have to take
5 a look and see if there are alternatives that would
6 have the same effect on your potential purpose without
7 having that disparate impact on vulnerable people.
8 Three different areas to think about, which
9 groups to assess. The protected classes include race,
10 gender, religion, and ethnicity. One of the things
11 that we encourage our members to think about is
12 expanding to vulnerable groups that are also at risk,
13 but are not explicitly protected by law, and which
14 purposes to assess. The law right now protects
15 eligibility decisions in employment, housing,
16 insurance, and credit.
17 But there may be other areas that are not
18 covered by existing laws where the decision-making is
19 consequential for people's lives and a company should
20 be thinking about whether or not to have the same kind
21 of disparate impact assessment in those contexts.
22 So there is a lot more to talk about. I am
23 delighted to be here at this panel. Thank you for
24 having me, and I look forward to the conversation that
25 follows.

1 (Applause.)
2 MS. CHOWDHURY: Thank you. I am the Global
3 Lead for Responsible AI at Accenture, and I am going
4 to be talking about understanding algorithmic bias,
5 particularly with a focus on consumer harms.
6 Much of our narrative today is about primary
7 harms. How do we expand and understand the
8 conversation about secondary harms and what are these
9 secondary consumer harms that we might want to think
10 about?
11 But, first, as a bit of background into our
12 practice, I have a colleague, Deb Santiago, sitting in
13 the audience today. We lead our responsible AI
14 practice at Accenture. We want to understand the
15 social, regulatory, and economic impact of this
16 technology from development to deployment. We do
17 provide solutions for clients who are very active in
18 the responsible AI community, including groups such as
19 the IEEE, World Economic Forum, World Society of the
20 Arts, et cetera. So we take not only a U.S.
21 perspective, but also a global perspective of
22 industry, government, and citizens.
23 So just to take a step back and think about
24 why we need ethics. This space is actually very, very
25 new and this panel is very representative of how very

1 new this space is. We have researchers developing
2 research at the same time that practitioners, such as
3 myself, are deploying these solutions to clients.
4 That is pretty rare. So our pipeline needs to be very
5 short, but at the same time, we need to be very, very
6 careful about what we are building and how we are
7 thinking about it.
8 Most of my time when I first started my job
9 in 2017 was spent building awareness. What is
10 responsible AI? The words we use today we did not
11 even have over a year ago. The way we refer to
12 things, the language that we are using, this evolution
13 of the space to think beyond technological tools to
14 now an evolved conversation about the human rights
15 impact, this is all happening at the pace at which you
16 are seeing it right now.
17 2018 was a year of action so Accenture was
18 first to market with a fairness tool. We alluded to
19 these concepts of fairness. Both my colleagues before
20 me alluded to these concepts of fairness. Our tool is
21 grounded in legal precedence so we have a disparate
22 impact component to our tool, and we specifically
23 think about the impact of the pipeline between the
24 legal and regulatory space to how we are applying this
25 in our solutions.

1 Finally, what we are thinking about moving
2 ahead is this concept of agency and accountability,
3 which is why I am here today, which is why the FTC is
4 considering artificial intelligence, ethical
5 frameworks, and how it impacts consumers. What we
6 have found from a technical perspective is we cannot
7 solve all the problems and maybe this is obvious to
8 the people in this room, but this is not obvious to
9 Silicon Valley. That we could not solve all the
10 problems by pushing buttons, writing code, and fixing
11 our data.
12 What we realized in the Amazon HR example
13 that Jimmy pointed out is a very good example. That
14 is actually, in my opinion, an example of good
15 governance. They tested a product, they innovated
16 safely, but they actually found that it was an
17 intractable human problem. Their hiring practices
18 were unfair. That is not a data solve. They tried
19 for years to make a data solve. But, ultimately, the
20 question becomes, well, Amazon, now that you have this
21 information, what will you do with it? That is where
22 the systems of agency and accountability come in.
23 Thinking on a more granular level, if an
24 individual algorithm has a negative outcome, then who
25 is responsible for identifying what that harm is and

1 adjusting and readjusting the harms. As citizens and
2 as consumers of this technology, who do I go to if the
3 Amazon recognition system falsely identifies me as a
4 pickpocketer? I know what to do if there is, for
5 example, a biased police officer. We have systems of
6 addressing and readdressing these problems however we
7 may feel about them. We do not have an infrastructure
8 of addressing and readdressing the harms that are done
9 to people by artificial intelligence.
10 So to think a bit about what is bias, Jimmy
11 did a really great job identifying from almost a
12 technologist perspective what is bias. We think of
13 bias as a quantifiable value. As a social scientist,
14 I would often call these experimental bias, so things
15 like sentiment analysis, things like imperfect data.
16 But really the takeaway here is that for us,
17 often when we think of bias, it is a measurable value
18 and often something you can fix if you just throw
19 enough data at it. If you fix your data, you clean
20 your data, you bootstrap your data, we will be able to
21 fix this bias. Or if we change our model, change some
22 parameter, we are endlessly tweaking and changing to
23 address this kind of bias.
24 However, when nontechnologists talk about
25 data, often we talk about the societal bias. And

1 these four harms listed were developed by the Future
2 of Privacy Forum and I think they encompass the kinds
3 of primary harms that we talk about today, economic
4 loss, loss of opportunity, social detriment, and loss
5 of liberty; things like the COMPAS algorithm, denying
6 people bail -- I am sorry, denying people parole
7 unfairly. So this is a loss of liberty.
8 But when we think about bias, we are also
9 often thinking about primary harms. So being
10 specifically denied a job when I am of a protected
11 class is something that is illegal. Now, if we could
12 define all of the harms neatly into those kinds of
13 buckets, frankly, we would not be holding this panel
14 today because existing law would be more than
15 sufficient to address all the harms that are happening
16 or at least the implementation of existing law.
17 Instead, I want us to think about secondary
18 harms, so this concept I am calling algorithmic
19 determinism. And one thing I want to point to as a
20 good example of algorithm determinism is the filter
21 bubble. Now, what is interesting is we have been
22 talking about the filter bubble for over a decade. We
23 have been living in the filter bubble for more than a
24 decade. The book, The Filter Bubble, was published in
25 2008.

1 So the question today is, does the filter
2 bubble lead to ideological polarization? And if you
3 are unfamiliar with the concept, a filter bubble is
4 when a recommendation system, an algorithm built by a
5 search engine provider or a media outlet is curating
6 data based on how you are reading information. So
7 what is the incentive of a media company? It is to
8 give you things that you will click on and read. But
9 what happens as a result is ideologically you start to
10 live in an information bubble. You have no idea or
11 concept of what other people are talking about that is
12 different from your notions and your ideas.
13 Why is this dangerous? The way these
14 algorithms will work often is they will increasingly
15 polarize you towards the opposite end of the people
16 moving away towards the center. And there is two
17 reasons this is dangerous. Number one is the obvious
18 one because I do not know is what is happening in the
19 world and I think that I am always right.
20 But I think the most dangerous one, number
21 two, is that if someone were to come to me as a human
22 being and say, I actually think a totally different
23 thing from you, I would actually just think they are
24 crazy as in you have no grounding, all the science
25 backs me because that is all I know and all I see, and

1 that inability to communicate on equal ground is
2 really dangerous.
3 But what I will add to this, this narrative
4 is important because it is not as if we as consumers
5 are battling this, we welcome this. Confirmation bias
6 is a very real thing. We love being right. We love
7 having our opinions affirmed and what happens here is
8 often we are battling our own inner biases. Our
9 desire to be right. We do not like it when we are
10 wrong. We do not like if somebody challenges us. So
11 we are not just battling an algorithm trying to guide
12 us in a particular way; we are also battling our own
13 nature.
14 So another example -- and this is an example
15 which starts to get into secondary harms, right.
16 There is nothing actually illegal about Netflix
17 targeting users by race. So why are we so upset about
18 it? Why do we think there is a problem with black
19 people being shown images of black people and women
20 being shown, you know, movies with a strong female
21 lead, which is often what I will get in my Netflix
22 queue. But we know that there is something wrong.
23 Otherwise, this would not headlining in The New York
24 Times.
25 And because, as I mentioned, we do not yet

1 have the language in the responsible AI community for
2 many of these things, I invite the term “algorithmic
3 determinism” to think through these secondary harms.
4 Why are we so worried about it? Because we are about
5 a world in which we only identify ourselves by a race,
6 we only identify with people who are of the same race,
7 who are only interested in media that looks exactly
8 like me all the time. What that does is reduce our
9 ability to be empathetic toward other people and other
10 people’s life situation.
11 So from a quantitative perspective,
12 algorithmic determinism is a measurement bias plus a
13 feedback loop. So a measurement bias ties into what
14 people like myself do which is literally the data
15 bias. And a feedback loop is something -- it is an
16 engineered loop where your output starts to influence
17 your input. If we think about artificial intelligence
18 as an algorithm that learns from its environment,
19 well, if I put something out there and I assume
20 something about the world and then by doing so I make
21 the thing happen and then I use that data to feedback
22 into my algorithm, I am creating a self-reinforcing
23 hypothesis.
24 So algorithm determinism starts to not only
25 make wrong assumptions -- that is only half of it.

1 The other half is it creates the world in which the
2 wrong assumptions are now true.
3 So measurement bias, as I mentioned, what
4 you think you are measuring is not what you are
5 actually measuring and a feedback loop is a structure
6 that causes an output to eventually influence its own
7 input.
8 So just in conclusion, I invite a
9 conversation around different types of bias. So what
10 does bias mean to different parties as technologists
11 and nontechnologists try to bridge a gap between our
12 lexicon? Let's make sure we are on the same page
13 about what we mean.
14 And second is that, as I mentioned, humbly
15 speaking as somebody in the responsible AI community,
16 we are still building our own lexicon, our own
17 language. Our language of harms needs to evolve to
18 embrace algorithmic determinism and the effects of
19 secondary harms. Agencies and bodies like the FTC,
20 who are dedicated to protecting consumers, can also be
21 involved in this conversation and thinking about not
22 just the primary harms, the direct harms to people
23 being denied services, but what are the long-term
24 impacts to society that may happen as a result of
25 algorithmic determinism.

1 Thank you.
2 (Applause.)
3 MR. WATTENBERG: All right. Thank you very
4 much. Thanks to the FTC for having me here. I am
5 delighted this conversation is taking place. And
6 thanks to the other panelists.
7 So I co-lead a group at Google called the
8 People + AI Research Initiative. Our goal is to make
9 human AI interaction better, to make it more
10 productive and fair. We take a broad view of this
11 mission. For one thing, we are interested in all
12 types of people, whether consumers, people who are
13 professionals, like doctors using AI, or engineers or
14 other developers of systems. We think it is important
15 to think about how all of these people work with AI.
16 We also produce a wide variety of work from
17 fundamental research that we write up and academic
18 publications, educational material, but we also do
19 engineering. We build tools and those tools are the
20 main subject of what I am going to talk about today.
21 So why are we building tools? Well, let me
22 take a step back and talk a little bit about Google's
23 AI principles. You can see them here. These are
24 principles that sort of guide us internally and
25 externally that we see as a kind of stake in the

1 ground. Some of these, in particular, I think
2 technology can actually help with. You know, we have
3 heard today that technology is not all of the
4 solution, but technology certainly has a role to play
5 in making things better.
6 In particular, as we seek to avoid bias or
7 avoid reinforcing existing bias, create safe and
8 accountable systems, and just uphold good standards of
9 excellence, tools can be very useful, and I want to
10 talk about a suite of tools that we have released to
11 the open source world. These all have a theme and the
12 theme is helping humans understand AI. For us, we
13 feel the route of -- sort of the best path to moving
14 forward is to increase our knowledge of what is going
15 on with AI systems. You know, it is important I think
16 both from an engineering perspective and to make sure
17 ethically that we are doing the right thing.
18 You hear a lot that people use the phrase
19 “black box” in talking about machine learning. And it
20 is not wrong in the sense that, you know, it can be
21 difficult to understand certain types of models. The
22 field is moving quickly. However, I think it is
23 inaccurate and there are often many ways that we can
24 actually get a handle on what is going on in systems
25 and then use that knowledge to make improvements.

1 One very important point I would like to
2 make is that people often talk about transparency as a
3 key value and transparency really has a lot of
4 different meanings here. It is not only as useful to
5 get full knowledge of a system. I mean, just to, you
6 know, give it a kind of silly example of like, you
7 know, if I wave my hand like this, you know, why did I
8 do this. One answer would involve every state of
9 every neuron in my brain, not very useful, or the
10 answer might be to take a rhetorical point, which is
11 useful.
12 Similarly, when you think about AI systems,
13 there are cases where an engineer might need a whole
14 lot of detail to debug a particular issue, but there
15 are cases where a consumer might be overwhelmed by a
16 lot of detail and might need just the type of
17 information they want to make a particular decision or
18 perhaps contest a decision.
19 Okay. So given that this type of knowledge
20 and understanding of AI systems is important, what can
21 we do to help with that? So one issue is to think
22 about the data that these systems have been trained
23 on. So as we have heard, training data is sort of a
24 key part of any machine-learning model. It really
25 determines the behavior. In fact, arguably, that is

1 the definition of machine learning is that the
2 training data does determine the behavior.
3 So, in order to understand what a system is
4 doing, it means we need to understand something about
5 the data very often. Now, this is hard because we are
6 dealing often with a lot of data, very complicated
7 data, and, generally speaking, people are not
8 incredibly good at sorting through data unless they
9 have a lot of expert training. Just looking at a
10 human table of numbers is overwhelming for almost
11 everyone.
12 But here is a place where technology can
13 help. One approach that my group takes to some of
14 these problems is with data visualization. So one
15 tool that we have released is called “Facets.” And
16 the idea here -- you can see sort of an animation up
17 here that shows this tool in action -- is that it lets
18 you slice and dice this data set in various ways. You
19 can look at quite a lot of data points. You can
20 divide them into groups; you can divide them into
21 subgroups.
22 One way to look at it using language we have
23 heard today is this is a tool for understanding
24 intersectionality, that we can actually see how
25 different groups interact with each other inside of

1 the data. And often using a tool like this, you can,
2 as a human, start to get a sense of what is going on,
3 what might be driving an issue with your data, what
4 might be potentially an issue that you have not seen
5 yet in behavior. So this is one very important way
6 that we can start to get at what is going on.
7 Okay. So data is one aspect. What about a
8 model itself? Very often, if you have a machine-
9 learning model that you are trying to analyze, you
10 want to ask it questions. You want to know things
11 like, okay, so I understand how it does on the
12 training data, what if I gave it something that was
13 completely different from anything in the training
14 data set, how would that affect things? Or say it is
15 a classifier and it classifies a data point in a
16 certain direction, you might say, what would change
17 that classification? You might want to fiddle with
18 particular aspects of that data point or ask what is
19 the most similar thing that was classified
20 differently.
21 So these are natural questions and I think
22 anyone working with machine learning is familiar with
23 this kind of thing. The problem is that they
24 typically require programming that requires
25 engineering time to do this. That means that

1 stakeholders, people who are not fluent in programming
2 languages may have a harder time getting answers to
3 these questions. So an approach that our group at
4 Google has taken is to create a tool that let's people
5 do this without coding. This is something we call the
6 "What-If Tool" and it is designed exactly to take a
7 machine-learning model in and then let you pose to it
8 hypothetical questions.
9 You can see sort of the animation, walking
10 you through a little bit of what is going on there.
11 It is built -- you know, Facets, that visualization we
12 just showed, is part of how this works. And it is
13 kind of a Swiss Army knife for understanding what is
14 going on in a model.
15 Now, there is something else. In addition
16 to looking at what is happening with an individual
17 data point, we can calculate more global statistics.
18 And this has a lot of helpful uses. One is for
19 thinking about fairness. One thing we can do is if
20 you define particular groups, then you can sort of
21 look at various group-based fairness measures. Now,
22 as we heard earlier, there are actually many different
23 mathematical measures of fairness. I think sorting
24 through these is an important issue for the community.
25 We do not take a position on this, but we do

1 offer people the option of saying, okay, I would like
2 to measure my system in various ways. We go one step
3 further then which is to say, if you have a threshold
4 base classifier, something very common, then we can do
5 a little optimization and say if it is not fair
6 according to this particular criterion, how would you
7 change the threshold to make it fair or as fair as
8 possible? So this gives you actual actionable
9 feedback that you could use with your system.
10 Now, again, I want to emphasize that as we
11 have heard so far, fairness is a very deeply
12 complicated sociotechnical issue and in no way do we
13 claim that just tweaking a threshold is going to fix
14 every problem. But it is something that can be an
15 important part of understanding a system and thinking
16 through ways that will lead to a solution.
17 I want to end with one other technology that
18 our group has developed and this is for looking at
19 neural networks. So 95 percent of the time that you
20 hear people talk about machine-learning systems being
21 black boxes, they are talking about what are called
22 deep neural networks. And the truth is that these
23 networks are complicated. You know, they are
24 typically specified by several very large matrices
25 filled with numbers that can look random at first

1 glance. So they can be difficult to analyze.
2 They are also often used on data sets that
3 themselves are difficult to understand. A classic
4 example would be image recognition. You know, suppose
5 you have a system that is designed to recognize
6 whether an image is a zebra or not. It is looking at
7 individual pixels and a lot of classical methods will
8 tell you things like, okay, did this particular pixel
9 make a difference to the classification? Did that
10 particular pixel make a difference? It is not super
11 useful looking at individual pixels. Instead, you
12 really want to look at something like, did stripes
13 makes a difference.
14 So the method that we used is something
15 called TCAV. It stands for testing with concept
16 activation vectors. This is introduced in a recent
17 paper by Been Kim and others. It is released as an
18 open source tool as well. What it does is it uses
19 machine learning to help you understand machine
20 learning. After something is trained, you can give it
21 examples of a concept you are interested in. For
22 example, for stripes, you might give it, you know,
23 say, 20 examples of striped rugs or shirts or
24 whatever. And then you can ask it questions. How
25 sensitive was that zebra classification to the concept

1 of stripes?
2 And so this is I think a very good example
3 of the type of translucency that is helpful. We are
4 not giving a researcher or a person looking at the
5 network the full matrix of every weight in the neural
6 network, but we are giving them information that is
7 useful at the level that they want in terms of a
8 concept that they are actually interested in.
9 So I would like to end there, but the point
10 I would like to emphasize is that there are many ways
11 in development we are making real progress in coming
12 up with ways to understand these systems. And I think
13 they no longer need to be considered black boxes.
14 (Applause.)
15 MS. LEE: Good afternoon, everyone. My name
16 is Erika Brown Lee, and I am at Mastercard. It is a
17 pleasure to be here, and when I say here I do not just
18 mean Howard University Law School, but participating
19 at the FTC's hearing on competition and consumer
20 protection.
21 As a former FTC person, I spent ten years at
22 the Commission in roles on the competition side and
23 the consumer protection side. So I appreciate the
24 opportunity to be able to participate in hearings that
25 are covering both sides of the Commission's mission.

1 Say that five times fast.
2 But before sharing my perspective with you
3 on AI, I thought I would turn back the clock a bit.
4 Not too much, but just for a few years, when you think
5 about -- and some of you in this room might actually
6 be familiar with AI from the concept of a movie that
7 was released sometime ago called "War Games." And
8 when you think about that movie, there was a computer
9 named Joshua who had to actually learn and self-teach
10 so that it would prevent nuclear war.
11 Well, that movie could have been made
12 credibly in 2018, but it was actually released back in
13 1982. So, of course, back then, artificial
14 intelligence was a lot more aspirational. But due in
15 part to the computational power -- the increase in
16 computational power you have heard from not only this
17 panel, but earlier in the day, and access to available
18 data, we now use artificial intelligence as part of
19 our daily lives. And the last panel talked about
20 examples of that, of the innovation behind AI powering
21 healthcare to detailed subway maps to computer vision.
22 But the agility of AI really presents these
23 opportunities for innovation. And at Mastercard, we
24 use artificial intelligence for fraud protection to
25 make our payment system safer and more secure for

1 cardholders. But as I think you have heard from my
2 colleagues on the panel, there are some opportunities
3 also for some structure around the discussion of
4 ethics in the deployment of AI.
5 So ethics is somewhat of a diffuse concept
6 just like fairness. It may mean different things to
7 different stakeholders, but several themes have
8 emerged to form a common set of principles. And I
9 wanted to cover a few of those principles today,
10 including transparency, accountability, and privacy by
11 design.
12 I will start with transparency because of
13 its role in building and maintaining consumer trust,
14 which is a key part of the ethics equation. Consumers
15 need to trust, need to have trust to be able to want
16 to share their data and have confidence in sharing
17 their data with entities. And so openness is a part
18 of the process for gaining and securing and
19 maintaining that trust and it can facilitate that
20 confidence.
21 But by openness, I am not referring to the
22 publication of algorithms. Martin just talked about
23 the deep neural networks or resource codes. From a
24 consumer perspective, I am not sure how meaningful
25 they would find them. A few months ago, Harvard

1 Business Review published an article about a case
2 study involving a Stanford professor, Clifford Nass,
3 who faced a student revolt. What happened? Well, the
4 students in his class claimed that the professor's
5 teaching assistants were grading the same type of
6 material in different ways. And so on their final
7 exams they were getting disparate grades.
8 It turns out they were right and the
9 professor agreed that there is a disparate outcome,
10 and so as a computer scientist, he designed a
11 technical fix and built a model to adjust the scores.
12 And in the spirit of transparency, he provided by
13 email the full algorithm to the students. But the
14 result was that the students were actually more angry
15 and there were more complaints. So it was hard to
16 reconcile this level of transparency.
17 So two years after the student protest, some
18 of the professors -- another professor's student
19 decided to do a study to explain what happened and in
20 that study the students were provided different levels
21 of transparency about the grades they received on an
22 essay. And it turned out that while medium
23 transparency increased trust significantly, high
24 transparency actually eroded the trust completely.
25 So the derived conclusion was that users did

1 not necessarily trust black boxes -- you have heard a
2 lot about those -- but that they did not really
3 necessarily need or want full transparency, but
4 actually enough information about the basic insights
5 and the factors driving the decisions that were based
6 on the algorithm.
7 But context matters. So the idea of
8 transparency varies depending on the context. And so
9 for example, if there is a smart washing machine, the
10 explanation of the decisions behind how to get your
11 clothes clean are quite different in need from
12 decisions about credit scoring or learning or lending,
13 for example. So there is a difference in terms of
14 context.
15 The other aspect of -- the other principle I
16 want to cover is accountability. And accountability
17 carries forward that level of trust and competence of
18 consumers, but there are several different levels of
19 accountability. On a macro level, accountability can
20 show how AI systems or models are ethically used to
21 create social value. At a more micro level,
22 accountability involves reviewing and assessing those
23 established objectives of an AI system.
24 And we talked about some of those or you
25 have heard some of those ways in which, from a

1 technical perspective you can accomplish that. But by
2 documenting the review and assessment, it can provide
3 a means of creating that feedback loop that can help
4 in understanding ongoing performance and identify some
5 of those anomalies and unintended -- perhaps
6 unintended consequences that Jimmy was talking about
7 earlier.
8 Accountability also provides oversight of
9 the technical administrative and administrative
10 controls. We are all familiar with audit, you know,
11 an audit, for example, of access controls. But given
12 the substantial increase of data that is collected by
13 an AI system, those technical controls become even
14 more important.
15 So the last principle or theme that I wanted
16 to talk about is privacy by design. An important part
17 of the exercise really of using an AI system is to
18 reconcile the tension between the protection of
19 individual privacy and the benefits from pursuing that
20 access to data that I was just talking about that AI
21 needs to be innovative and to work efficiently.
22 Privacy by design can reconcile those two
23 competing interests. So by imbedding privacy into all
24 of the stages of development -- so from that I mean
25 from design -- well, really from ideation then design,

1 build, testing, deployment, privacy can actually be
2 used as a strategic asset. So for example, the
3 concept in privacy -- one of the key concepts is
4 minimization which calls for limiting the amount of
5 data that is collected. That may at first seem to be
6 contrary to how AI systems work and what I was just
7 talking about in terms of availability of data.
8 Well, at a certain point, an AI system may
9 actually not benefit from the increased value or the
10 increased amount of data; in other words, if it is not
11 necessarily improving the success or efficiency of the
12 result. And so limiting data may improve efficiency.
13 Or it may be that data becomes less relevant. And so
14 over time that may also encourage minimization.
15 Privacy by design we heard a little about
16 that, the legal requirements. Data flows across
17 borders. So even though we are contemplating more of
18 the U.S. perspective here, it is important to consider
19 from a global perspective as well because other
20 jurisdictions have, in fact, restricted and added
21 additional requirements with regard to transparency or
22 consent from the individual to use their data.
23 And a privacy impact assessment can be used
24 to identify those potential risks and harms to
25 individual privacy and strategies for managing those

1 risks. The idea is that if you incorporate privacy,
2 in particular -- and again it is not sort of a one
3 size fits all, but incorporated appropriately, it can
4 enhance the AI profile.
5 One other point I wanted to make before
6 concluding is just about data literacy, which is
7 something that goes hand in hand with privacy and it
8 is part of the broad theme of accountability because
9 data literacy extends from the ideation stage and with
10 the computer scientists and coders all the way through
11 launch of a product.
12 But I will conclude by saying that as we go
13 forward, it is important to have standards that are
14 consistent standards that are flexible and inoperable
15 not just in the U.S., but globally and that ensure
16 meaningful protections of privacy.
17 So I will stop there and turn it over to
18 Naomi.
19 (Applause.)
20 MS. LEFKOVITZ: Okay, thank you. And thank
21 you for having me here today. It is a pleasure.
22 So I am going to talk a little bit about
23 sort of the research and standard space and also a
24 little bit about where NIST is trying to contribute to
25 some foundational concepts and privacy risk management

1 and engineering and see how they might apply in the AI
2 space.
3 So at NIST today, we have about -- more than
4 50 projects that are either contemplated or underway
5 in artificial intelligence and machine learning, and
6 many of these are focused on exploring fundamental
7 questions related to measurement and quantification.
8 And I do not have even barely the time -- I do not
9 have any time, right, in ten minutes to talk about all
10 of these projects. So I really just want to make sort
11 of a key point that you have sort of heard that we
12 have to understand what kind of assurance we can get
13 about the correct operations of AI systems. And I
14 think you have already heard today that even
15 "correct," right, is sort of a complicated concept
16 and has different view points on that.
17 But at a bare minimum, right, if we want to
18 have AI systems adhere to ethical frameworks, we
19 really need to understand what that correct operation
20 means in that context. Otherwise, we really do not
21 know if they are going to adhere to them.
22 So the next set of slides I am going to run
23 through, I am not going to talk to these individually.
24 What I really just want to share with you and I know
25 that these -- I understand these slides will be posted

1 so that you can look at this and get a better sense if
2 you are really interested into where the sort of scope
3 of work is going around various standards.
4 And so the second point I want to make is
5 that these are not actually finished standards.
6 Nothing that I am going to show you in the next set of
7 slides -- you will see study, you will see all kinds
8 of terms, but none of them are completed standards.
9 This is beginning work.
10 Why do standards matter? Let me give one
11 example, not in the AI space. So we were working in
12 the identity federation space and wanted to see more
13 privacy-enhancing technologies integrated. And what
14 we quickly discovered was that the underlying
15 protocols on which sort of identity federation is
16 running had never contemplated some of the integration
17 that we want to do and literally in terms of sort of
18 like, hey, we want to put this key exchange in here
19 for this privacy-enhancing cryptographic technique and
20 there is no field for that in the protocol. People do
21 not like it when you break protocols, when you break
22 standards because the point is everyone is trying to
23 build their systems to use these standards so that
24 everybody can communicate interoperably.
25 And so it is actually very important to

1 build in some of these -- what you want out of the
2 system either from ethics or privacy into these
3 standards or be thinking about that because if they
4 get designed, if these sort of underlying standards
5 get designed without that, it is very hard to go back.
6 You can go back and redo the standard, but it is very
7 hard to get your additional technologies sort of
8 retrofitted in there.
9 And the other point that I want to sort of
10 make is on some of the challenges in this standard
11 space. So you can see that there are these different
12 types of standards. Some of them are very specific,
13 like a standard for ethically-driven nudging for
14 robotic intelligence and autonomous systems. But you
15 see over here in ISO they have all these different
16 working groups -- that is what WG stands for -- and
17 you can see -- so, for example, SG 1, there is that
18 computational approaches and characteristics of
19 artificial intelligence systems. If you are not
20 thinking about sort of those ethical characteristics
21 and people in there are not thinking about it, the
22 ones who are actually building that standard, it is
23 not going to done.
24 So it really takes engagement and you can
25 see there are these multiple groups and they are all

1 working on these different areas. And you try to have
2 liaisons, but it is challenging and something to be
3 aware of and why NIST encourages everyone who can to
4 get engaged in the standards development so they get
5 developed the way we think they should. So I am going
6 to move on and you can look at these.
7 Now, I am going to talk a little bit about
8 some of the NIST work. So we introduced some
9 concepts, some constructs around privacy engineering
10 and risk management because we saw some of the same
11 issues that are coming up. What do you do with
12 principles that are sort of this high level and how do
13 you deal with them down at the implementation stage?
14 And so you know, I will admittedly say that
15 we are using the term “privacy,” but it is an
16 imperfect word and you will see that I think we cover
17 a lot of the things that people are talking which
18 might, in some people’s minds, go beyond the concept
19 of what they think of as privacy.
20 The main point here is that first we began
21 to have -- you know, we have some of the same issues
22 like lexicon and language, what are we talking about.
23 Mainly people think that, okay, if I have protected
24 data, I have managed privacy. But, of course, there
25 is more to that. Sometimes we talk about an example

1 with the smart grid, right. So the reasons that some
2 communities were objecting to smart meters was not so
3 much because the utilities could not keep the
4 information secure, but because the smart meters were
5 collecting such detailed information that inferences
6 could be made about their behavior inside their home.
7 So how do we manage some of those? Well, in
8 security, right, when we want to understand how do we
9 deal with implementation, right, I mean, how do we go
10 from principles and how do we apply them. We tend to
11 use a security risk model. And so here I think
12 everybody knows there is -- you know, what is the
13 likelihood that a threat can exploit a vulnerability
14 and what is the impact. But how do we apply that in
15 the smart grid space? What is the unauthorized
16 activity that is happening? What is the threat? The
17 smart meter?
18 So we had some concerns that that was not
19 necessarily the greatest model for the full scope of
20 privacy risk. And so what we said was what is the
21 adverse event and what are some of the things that you
22 have been hearing about. We have heard it in
23 different terms, secondary harms, primary harms. We
24 went with the term “problems” to sort of distinguish
25 from things that might be legally cognizable versus

1 things that are going to be troublesome for people and
2 that organizations may want to manage regardless of
3 whether there is a legal cost to it or not.
4 So you can see that there is a whole variety
5 of problems. These are nonexhaustive and you can put
6 sort of anything in there that you want that people
7 can experience. And that allows us to have this model
8 where we can say what is the likelihood that any kind
9 of processing of data, any particular operation could
10 create some kind of problem for individuals and what
11 would be the impact. And that is really the heart,
12 right, of where you go from principles to, you know,
13 what people -- my panelists have been talking about
14 which is like, well, how do you change the context?
15 How do you understand how much transparency to have,
16 right?
17 Well, we can think about sort of the impact
18 and we think about, hey, what do I want this AI to be
19 doing and how do we want it to impact or not impact
20 individuals? This is where a risk model and risk
21 management processes can come into play.
22 The final thing I would briefly mention is
23 the construct that we introduced in our NIST report,
24 is the concept of privacy engineering objectives. And
25 these are essentially additive to the security

1 objectives, confidentiality, integrity, and
2 availability. And so I think you have heard some of
3 the challenges around things like transparency, they
4 can be interpreted very differently. And so, for
5 example, we can elevate that into, as an objective, in
6 terms of what kinds of properties do we want our
7 systems to support, we can say, well, we would like to
8 enable reliable assumptions about processing.
9 And if we extend that to AI, we could extend
10 that to AI behavior. So we do not necessarily need to
11 know every detail, but we would like to have some
12 reliable assumptions. How much manageability, right,
13 or intervention, right. If I am driving a car, I can
14 make a choice to hit a squirrel or save my child,
15 right. So I can make those choices and I will take
16 the consequences for that. But what about the AI? Do
17 I have any ability to intervene in whatever
18 programming and decision-making it is making about
19 that?
20 And then disassociability is really about
21 being able to disassociate information from
22 individuals and device.
23 So with that, I will end. Thank you.
24 (Applause.)
25 MR. TRILLING: Thank you to each of our

1 panelists for the excellent presentations. To start
2 things off for the discussion portion of the panel, I
3 want to remind our panelists to please turn your name
4 cards to the side if you want to weigh in.
5 I want to start off with a fairly broad
6 question. So over the course of the day, we have
7 heard references to a number of different ethics
8 concerns and other constructs related to ethics. For
9 example, we have heard about transparency,
10 accountability, privacy, bias, fairness. My question
11 is: Are the ethical concerns raised by artificial
12 intelligence different from the ethical concerns that
13 are raised by traditional computer programming
14 techniques or by human decision-making? And if so,
15 how and why?
16 James, do you want to start? Jimmy?
17 MR. FOULDS: Okay. So first, I would say
18 scale is a big difference. Now, so you can build an
19 AI system and then deploy it on millions of people
20 with a few clicks of a button. So just the share
21 scale of potential impact on people, I think that is a
22 big one.
23 Another one is kind of transparency is
24 different versus human decision-making. In some
25 sense, everything is there in the computer, right?

1 You have a model, or an algorithm that is making
2 decisions and it is all digitally encoded. But it can
3 be difficult to understand what that means or what it
4 is doing.
5 So Martin was speaking to ways we could try
6 to unpack that, but it is a difficult challenge,
7 whereas as Rumman mentioned if you have a human, you
8 can go and ask them why they made a decision, but we
9 may not be able to do that for algorithms.
10 MR. TRILLING: Rumman, do you want to go
11 next, please?
12 MS. CHOWDHURY: Sure. So to echo Jimmy a
13 little bit, I have what I call the three Is, AI is
14 immediate, impactful, and invisible. And what that
15 means is when you deploy an artificial intelligence
16 system, it impacts as wide of an audience base as you
17 have. So you think of a social media company making a
18 change to its algorithm to show you media. It happens
19 right away. There is not oversimplifying the
20 engineering process here, but there is not like this
21 wait period where you ramp up.
22 The impact -- and this is what Jimmy was
23 talking about, you touch people's lives in very
24 meaningful ways with artificial intelligence. And
25 this is different from traditional computer systems

1 and traditional methods of thinking about computation.
2 As opposed to systems like maybe a car or a
3 television, which is tangentially related to our
4 lives, as much as I may love watching Netflix, it is
5 technically tangentially related to my life, the
6 algorithms that influence my life are things that
7 actually are literally impacting my life choices.
8 And, finally, they are invisible, so this
9 notion of a lack of transparency. But also the fact
10 that I do not always know when there is an algorithm
11 impacting my experience. I am not sure if I am being
12 shown something because it has been hard-coded or
13 selected for me because there is an algorithm. Now,
14 if you think about the notion of bots on social media,
15 those are algorithms posing as human beings. I may
16 think I am being given media or told some information,
17 but I am actually not. It is being created by an
18 algorithm. So thinking about the difference between
19 AI and traditional computing, specifically with the
20 three Is and importantly about the pervasiveness.
21 MR. TRILLING: Mark, did you have something
22 to add?
23 MR. MACCARTHY: Thanks. Let me emphasize
24 the continuity rather than the discontinuities. Many
25 of the same issues that we run across in the older

1 regression analyses models, the credit scores, the
2 recidivism scores that are so controversial right now,
3 provide very good models for how we should think about
4 the ethical issues involved in machine learning and
5 other AI systems.
6 I think the techniques of explainability, of
7 providing reasons, identifying the major factors that
8 credit scoring companies have been involved in for a
9 generation are useful lessons for AI algorithms as
10 well. You get into a slightly different set of issues
11 when you come to autonomous systems, where the
12 activity really can take place without human
13 intervention. Autonomous weapons where you say, pick
14 your mission and then go execute it, without human
15 intervention, those raise ethical issues that are
16 quite different than standard regression analysis and
17 they deserve different thinking. Same with autonomous
18 cars, to the extent that they are making decisions
19 about what to do on the road without human
20 intervention, those questions really raise some new
21 issues.
22 But for the most part, in the issues that we
23 deal with on an everyday basis right now, the new
24 systems really are largely similar to the older
25 systems, and many of the principles and many of the

1 techniques for thinking about these problems have been
2 developed for the earlier algorithms and can be
3 applied to the new cases as well.
4 MR. TRILLING: Martin?
5 MR. WATTENBERG: Yeah, I just want to add
6 that I think the focus on ethics is actually really
7 beneficial and is helping us even understand existing
8 systems better and what was good about them. So one
9 example that came up earlier is this idea that if you
10 take a human decision-making system and automate it,
11 you might lose the chance for contestability if you do
12 that in a careless way.
13 And I think what that is telling us is the
14 key issue was the contestability. It is less about
15 automation and not automation and more about what we
16 want as a society around that process. And I think
17 that is an important thing to keep in mind as we
18 think through these issues. Often, we discover
19 thinking about ethics in the context of AI we have
20 clarified our thinking about -- on AI systems, as
21 well.
22 MS. GOLDMAN: So I would like to ask a
23 question that is related to the last one in terms of
24 comparing AI to other more traditional methods of
25 analysis. And we have heard a lot of different

1 frameworks and principles for AI, such as the
2 fairness, accountability and transparency, Belmont
3 principles, SIIA, IEEE policy standards so. So there
4 are a whole lot of frameworks. And by thinking about
5 these different frameworks and applying them to AI,
6 are we holding them to different standards than would
7 be applied to human or other traditional decision-
8 making?
9 And, also, perhaps more conflicts and
10 case-by-case question, but how can compliance with
11 these ethical frameworks or principles be measured and
12 by whom?
13 Maybe we will just go down the line again.
14 James, would you like to start?
15 MR. FOULDS: So, first, I want to point out
16 that AI systems are engineered, right? They are
17 created. Even though they are run by mysterious
18 algorithms, they are generally put together by a team
19 of humans who work for a company and who will analyze
20 the performance of these systems and measure what they
21 are doing and decide if it is satisfactory. And so to
22 that extent, these systems are actually not that
23 different from other complex systems, such as the
24 creation of automobiles. So my view is that we should
25 hold them to similar standards to other complex

1 engineered systems like creating automobiles or
2 airplanes or spaceships, and so on.
3 In terms of how to measure these things, so
4 the machine-learning community has put together a
5 large number of definitions of fairness and so on. So
6 these are definitely tools that we could try to use to
7 measure if these methods are fair or not and then we
8 have to probably have a debate about which of them we
9 give the most weight to.
10 MS. GOLDMAN: Thank you.
11 Mark?
12 MR. MACCARTHY: Let me agree with the point
13 that there is a similar set of standards that apply to
14 AI and non-AI systems. I think the principles that I
15 cited are largely usable in many, many different
16 contexts. But that brings me to the measurement
17 question and I do not think there is a good way to
18 measure compliance with principles at that level of
19 abstraction. All of the key issues really are going
20 to be -- wind up being faced when you get to the level
21 of application. And there, I think measurement is the
22 wrong concept because it sounds like if you just add
23 and subtract enough, you will come up with an equation
24 that gives you the right answer.
25 In fact, these are very, very complicated

1 and difficult ethical question. It is not to say
2 there is no right answer, but it may be the kind of
3 answer that emerges from discussion, debate and
4 reflection on what we want as a society, rather than
5 measuring something and coming up with the right
6 answer.
7 To go back to the concepts of fairness that
8 were developed before, the computer science community
9 knows perfectly well that they are trying to provide
10 sort of computer science analogs of very basic, legal
11 philosophical and ethical concepts and they break into
12 two big parts, group fairness versus individual
13 fairness. And people differ in a large part whether
14 they think fairness is a matter of accuracy and
15 classification and that is it, or they think fairness
16 is a matter of protecting the interests of vulnerable
17 groups, including groups that have been historically
18 disadvantaged.
19 You get very, very different conceptions of
20 what the discrimination laws are all about, if you
21 take one of those two different points of view, and
22 then you develop very, very different computer
23 measurements of whether you have satisfied those
24 objectives once you bring it down to the level of
25 measurement. But the key concepts are fundamentally

1 ethical, philosophical, and legal. And they are not
2 concepts that are native to computer science.
3 MS. LEE: Okay, yeah, I think that the
4 question is very interesting because it really poses
5 something that as a community we need to think
6 through, in terms of whether -- you know, how ethics
7 plays out in decisions for AI.
8 There was a commentary from a German
9 parliamentarian when he was asked about the trolley
10 problem about what the result would be if a trolley is
11 going -- for those of you who do not know, if a
12 trolley continues straight and does nothing, then it
13 results in the deaths of everyone. But then if it is
14 diverted then, you know, some people die and others do
15 not, so sort of that ethical dilemma. And the
16 response was, well, whether it is a human making a
17 decision or an algorithm making the decision, there is
18 still a tragic result.
19 So from a human perspective it is just -- it
20 is going to be a split second determination that no
21 one really has time to think about. So you could
22 deploy that almost from a randomness perspective for
23 an algorithm and end up getting the same result. But
24 the creepiness of it comes from that transparency. So
25 how is it -- how is that decision being made? So my

1 panelists have talked about, it comes up a lot more
2 when the impact -- the higher the impact to the
3 individual. And so I do think it flows back to that
4 level of transparency.
5 But whether it is an AI system or not,
6 levels of transparency and the requirement to provide
7 additional information behind decision-making is long
8 embedded in U.S. law. And so I do not know that it
9 necessarily makes a difference whether it is an AI
10 system or not. To me, it comes down to the impact.
11 MS. LEFKOVITZ: So I guess I would say that
12 there are sort of different levels of measurement.
13 And part of that has to do with like what are you
14 looking for, right? So I think that has been
15 underlying a lot of the presentations today. And so
16 one reason that we went in the direction of privacy
17 engineering objectives was because of the fair
18 information practice principles are hard to sort of
19 measure. But you can measure what a reliable
20 assumption is, right? You can actually test that.
21 And so that is one of the reasons why I
22 think the confidentiality, integrity, and availability
23 have been successful as security objectives because
24 they break these things down into pieces that you can
25 then assess. So I think that is part of this

1 conversation today and that we will go on is figuring
2 out what are our objectives and how are we sort of
3 managing risk. What are we looking for? Then we can
4 know what we can measure.
5 MR. TRILLING: Are there ethical issues that
6 people are raising in relation to artificial
7 intelligence that may be misplaced? And if so, what
8 are some examples?
9 MR. MACCARTHY: I think the whole notion
10 that artificially intelligent systems will develop
11 consciousness and agency I think is so speculative
12 that it is not a real problem. Yet, is it the kind of
13 thing that absorbs a lot of time and attention, far
14 more than it really deserves, considering that there
15 are real problems associated with these systems that
16 need to be fully addressed.
17 MR. TRILLING: Rumman?
18 MS. CHOWDHURY: So I used to start all of my
19 talks by saying there are three things I do not talk
20 about, terminator, hell, and Silicon Valley
21 entrepreneurs saving the world.
22 (Laughter.)
23 MS. CHOWDHURY: So I would just add that to
24 the mix.
25 (Laughter.)

1 MS. CHOWDHURY: But I would also say that
2 often we over anthropomorphize artificial
3 intelligence. There is -- as humans, we like to
4 impose human features on things. And you think about
5 being a child and your favorite toy, which I may have
6 been a bear, but you gave that bear a name and it had
7 a personality, right, or you had an imaginary friend.
8 That is what we, even as adults, we humans like to do.
9 So one thing that particularly concerns me
10 is a sense of over-responsibility of the algorithm for
11 the negative outputs, a term I call "moral
12 outsourcing," where by anthropomorphizing the AI and
13 deflecting or pushing all the responsibility on the
14 artificial intelligence, by writing this narrative
15 that it is alive, it is making decisions, et cetera,
16 it has free will, we are removing the responsibility
17 from human beings and we are scaring ourselves away
18 from the narrative and from the ability to fix these
19 very human problems.
20 MR. TRILLING: Martin?
21 MR. WATTENBERG: Yeah, I think echoing what
22 you have heard, I would say it is not possible to
23 over-hype ethics. I think ethics is critical and this
24 focus is really, really good. It may be possible to
25 over-hype AI as we have heard. I think it is a tool.

1 It is an important tool and a very exciting one. But
2 in the end, it is a technology like many others we
3 have dealt with and I think we should deal with it in
4 the same way as we have dealt with other technologies.
5 MS. GOLDMAN: So this morning in Michael
6 Kerns' presentation, we heard some things about
7 tradeoffs between fairness and accuracy and even
8 tradeoffs between different types of fairness. So I
9 wanted to get this panel's take on those types of
10 tradeoffs and also, what are the considerations that
11 should govern the design of a system in which accuracy
12 and fairness are at issue?
13 MS. LEE: we clearly all have very strong
14 opinions.
15 MR. FOULDS: So, yes, there are definitely
16 tradeoffs between accuracy and fairness. Of course,
17 it depends how you define fairness. So there are some
18 definitions of fairness which only consider accuracy
19 as being a good thing. But there are other notions
20 more related to equality or parity where there is a
21 clear tradeoff between fairness and accuracy. So my
22 take on this is an accurate algorithm is not
23 necessarily a fair one because we need to distinguish
24 between the predictive task of classification or
25 making some prediction, assigning an outcome to a

1 person that makes a prediction versus how that is
2 going to be used, which is an economic question, what
3 is the impact of when I used this to make decisions on
4 people's lives, what is it going to do to them? What
5 is the effect on them and on society?
6 So an example that I like to use is college
7 admissions. So suppose you would like to use a
8 classifier, a machine-learning algorithm to determine
9 whether to admit people to a college. So you could
10 try to predict their GPA.
11 But we all know that we have a leaky
12 pipeline in STEM and in probably every field and that
13 can be impacted by unfair factors in society, like if
14 you are poor or marginalized, you are more likely to
15 get sick, you are more likely to have a mental
16 illness, you are more likely to have family members
17 who get sick, you may be far away from healthcare
18 where you live. So you are more likely to have your
19 grade harmed and drop out. So if you just try to
20 predict GPA and use that to determine admissions, then
21 your accurate classifier may not be a fair one.
22 MS. CHOWDHURY: So the way I think a lot of
23 us are inviting more granularity around the term
24 "fairness," invites more granularity around the term
25 "accuracy." So this is another one of those examples

1 of technologists and nontechnologists talking past
2 each other. Accuracy means something very, very
3 specific to us. It is a quantifiable value. Again,
4 when we are explaining machine learning -- supervised
5 machine learning as having your output, your accuracy
6 is just a measure of how often your testing data was
7 correct.
8 We take our data. We put it into two piles.
9 We train it on one algorithm and we check our homework
10 on the other. That is our measurement of accuracy.
11 Now, is that a measurement of accuracy we believe in
12 in the real world? Maybe, maybe not. So one might
13 say that sure, minorities underperform, does that mean
14 that they systematically underperform? That it is the
15 action of being of a particular race that makes you
16 underperform? No, we know that is not true. And this
17 is why we are concerned about proxy variables.
18 Another thing I am doing additional research
19 in, particularly in algorithm determinism, is this
20 concept of mutability and immutability of variables.
21 Algorithms do not know the difference between things
22 that we can change and things that we cannot change. I
23 cannot change my age; I cannot change my biometrics.
24 There are things about myself I can change,
25 maybe my educational attainment, my weight, my hair

1 color. But an algorithm does not know the difference
2 between two. So when we think about things like
3 accuracy, how much are we imposing that accuracy as
4 this objective truth or this objective world order and
5 how is that related to systems of fairness and
6 unfairness in our society?
7 MR. MACCARTHY: So fairness and accuracy.
8 Let me go back to the Netflix example that you raised
9 earlier. So accuracy, if a company is trying to
10 assess accurately the taste of people in movies, there
11 is a good chance you are going to get racial
12 differences among groups. It turns out people's
13 tastes differ by race.
14 Now, should you try to fix this? Is there
15 some unfairness involved in that? Well, you could
16 move away from accuracy towards a kind of group
17 equality. And your reasoning might be, well, you want
18 people to have a diversity of experience, maybe they
19 will see something that is not part of their prior
20 taste and they will learn a little bit more about the
21 way other people live. But the cost might be that
22 there would be a mismatch between the recommendations
23 and people's current taste.
24 So there is a tradeoff there. People have
25 to think about which one they want as a matter of what

1 we want our society to be like. But it is very
2 similar to what is going on in the recidivism scores.
3 But what this illustrates is that the way we make that
4 tradeoff and the importance that we ascribe to that
5 tradeoff differs by context. In the context of the
6 Netflix example and recommendations for movies, there
7 is one set of considerations.
8 But in the recidivism situation, there are a
9 whole bunch of different circumstances but a very
10 similar sort of structure. If you assess people's
11 likelihood of re-offending, it is going to turn out
12 that you are going to get racial differences. People
13 re-offend at different rates depending on their group
14 membership.
15 Now, should you fix this? There are a
16 couple of very strong reasons for thinking that you
17 circulate. One is that racial bias is endemic in the
18 criminal justice system and it is high time we do
19 something about it. The other is that in the criminal
20 justice system, one of the principles we kind of live
21 by is to protect the innocent. You know, we do not
22 want to catch the guilty so much as protect the
23 innocent. So for both of these reasons you might want
24 to move away from just trying to get as accurate a
25 predictor as you possibly can.

1 And you can do that by using one of these
2 other concepts of fairness. Group fairness, you can,
3 for example, equalize group error rates. The problem
4 is if you do that, you lose something called
5 predictive parity in the algorithms. And you raise
6 all sorts of complicated legal, philosophical, and
7 ethical questions involving due process,
8 constitutional questions, all of the difficulties
9 about affirmative action are things we have to start
10 to deal with. There is a cost as well in terms of
11 greater risk to public safety by taking that
12 particular direction.
13 Now, that is where you find the real ethical
14 issues, right. In that kind of tradeoff, you have to
15 talk about it in the concrete context of some
16 particular practice like criminal justice in order to
17 really get your teeth into the ethical problems. It
18 is not going to be solved and we are not going to make
19 process at the level of debating abstract principles.
20 You really have to look at those concrete cases to
21 understand how to make the tradeoffs.
22 MR. WATTENBERG: I would like to sort of add
23 a kind of practical note to this, which is that I
24 think theoretically you can point to situations where
25 there are real tradeoffs. But practically speaking in

1 my experience, when you have a system, you identify
2 some way that it is unfair and then find a way to fix
3 it. It actually gets better overall. And just to
4 take an example, one of the most common reasons for a
5 system not to be fair is that it has been trained on
6 the wrong data that is not representative of what is
7 happening in the real world that it is being served
8 on. And when you get better data, it is just a
9 blanket improvement or nothing gets worse overall.
10 That is just a good thing.
11 So in many cases, fairness is just a symptom
12 of other underlying problems and so I do not think
13 that we should assume there is always a tradeoff
14 between fairness and accuracy.
15 MS. CHOWDHURY: Sorry to step in, but
16 anecdotally, I have a similar example with our
17 Accenture fairness tool. When we were using a credit
18 risk modeling algorithm to determine whether or not a
19 system was fair or unfair by particular metrics,
20 disparate impact, predictive parity, when we actually
21 equalized for predictive parity by gender, we actually
22 found our accuracy rate improved. It improved because
23 we opened up credit opportunities to people who would
24 previously have been denied. So I absolutely agree
25 with you that it is not always a foregone conclusion

1 that fairness and accuracy are a tradeoff.
2 MR. FOULDS: I have seen a similar situation
3 where overfitting is the problem. So you have a model
4 that is too powerful, that fits too closely to the
5 data that can harm both accuracy and fairness, and I
6 have seen that happen.
7 MR. TRILLING: Naomi, did you want to weigh
8 in quickly before we move on to an audience question?
9 MS. LEFKOVITZ: Yeah, I just wanted to add,
10 I mean, this is why we came up with a privacy risk
11 model, right, because when you are in a tradeoff
12 space, it helps to have a frame of analysis. So in
13 that contextual space, you can understand, well, you
14 know, what is the impact that this measurement of
15 accuracy is having and how is that impacting or
16 creating problems for individuals? And then can you
17 begin to make decisions and find the solutions that
18 sort of both optimize your accuracy and also minimize
19 those adverse consequences.
20 MR. TRILLING: One of our audience members
21 has asked, what are the main sources of data that are
22 being used to develop algorithms, and if personal data
23 are a source, how are subjects informed? And I want
24 to relate that to a second audience question, which is
25 if the data are corrupt, is the fault left to data

1 scientists, programmers or someone else and who is
2 responsible for fixing the data?
3 MS. CHOWDHURY: I think those are incredibly
4 important questions. So just getting at the concept
5 of data consent, I think there is also an issue here
6 where there is a misunderstanding in the public about
7 what it means to give consent to data and what that
8 relationship with people and data are. So I am going
9 to sort of answer the question, but maybe take the
10 conversation to a little bit different place.
11 Most people understand a relationship with
12 algorithms and data or data scientists and data to be
13 similar to when you would give your email address to
14 get 10 percent off at some clothing retailer and then
15 they occasionally send you spammy emails. It is a
16 very direct relation. It is purely transactional.
17 And I know the analogy is data is the new oil. But
18 instead I think of data as a new periodic table. Why?
19 Because I can take the same element, hydrogen, and I
20 can use it to make water, something that gives us
21 life, or the hydrogen bomb, right, something that can
22 cause massive amounts of pain and destruction.
23 And data is very, very similar. What we do
24 not realize is seemingly innocuous data can be used in
25 many different ways. You may not care if a company is

1 picking up the number of steps you walk per day. But
2 when that may influence your insurance premium, you
3 will definitely care.
4 The problem with getting consent is that we
5 are not even shown what we are giving consent to
6 because the companies which we are giving consent
7 do not always know how they are going to use them.
8 And, also, are we giving data consent in perpetuity?
9 What if three years from now that is a very
10 viable algorithm where the number of steps I
11 walk per day cross by, you know, other seemingly
12 innocuous pieces of information, plus the IoT from the
13 publicly available cameras that are available in every
14 smart city will then be used to actually measure my
15 degree of health and, therefore, impact my insurance
16 premium.
17 What rights -- when I agreed to share my
18 number of steps, that algorithm maybe did not exist.
19 Now that it exists five years later, what rights do I
20 have over it? And these are the kinds of question
21 that we are trying to understand and grapple with and
22 that requires a very fundamental reworking of our
23 relationship as human beings with data.
24 The other thing I would point out within the
25 consent is we cannot -- even if we take back our

1 information or data or stop sharing, the historical
2 information we have given, we do not have rights over
3 that information. So what must we think about in
4 terms of data we have already provided or we have no
5 control over what we are providing if we are in
6 public, for example?
7 MR. TRILLING: Erika?
8 MS. LEE: So I agree with that. I think
9 that the question is such a good one about consent and
10 consumer control over data. It is hard to sort of
11 place and do the chain of activities that can be
12 undertaken once data is ingested. One of the things,
13 as I mentioned earlier, is sort of trying to do a risk
14 assessment, Naomi has talked about this, too, it can
15 be done through a privacy impact assessment and trying
16 to at least identify what the risks are.
17 One of the mitigation strategies that can
18 partially address the question is sort of
19 anonymization techniques or encryption techniques, but
20 anonymization, in particular, where you are separating
21 the identity of the individual from that data. So to
22 the extent that data can be anonymized, may be a way
23 to use the data -- somebody I think earlier talked
24 about, in addition, differential privacy where you are
25 sort of introducing noise to the data, so it does not

1 affect the integrity and the ability to use the data,
2 but still protects that information.
3 There are encryption -- also encryption
4 tools like the homomorphic encryption is just an
5 example, but there are strategies that potentially can
6 be deployed to still allow use of the data without
7 sharing or transferring some of that highly personal
8 data.
9 MR. MACCARTHY: So one last very quick
10 comment. All the difficulties of getting consent that
11 we have been talking about, I think that is one reason
12 why the NIST framework that Naomi was talking about
13 where the way of thinking is identify a harm that is a
14 possible harm and then assess the risk of that harm
15 and then take steps to mitigate it, that approach,
16 which puts a lot more of the burden on the data
17 controller than on the individual data subject, may be
18 a very productive way forward.
19 MR. TRILLING: So the bad news is we are out
20 of time. But the good news is that our next panel,
21 after we have a 15-minute break, I think will be in a
22 good position to pick up the discussion that we have
23 covered on this panel. So please join Karen and me in
24 thanking our panelists for a great discussion.
25 (Applause.)

1 MR. TRILLING: And we will return at 3:15.
2 (End of Panel.)
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 CONSUMER PROTECTION IMPLICATIONS OF ALGORITHMS,
2 ARTIFICIAL INTELLIGENCE, AND PREDICTIVE ANALYSIS
3 MS. GEORGE: Good afternoon, everyone. And
4 thank you for sticking around for the last panel of a
5 very full, exciting and informative day. Hopefully,
6 we can keep you engaged through this last panel.
7 My name is Tiffany George. I am an attorney
8 at the Federal Trade Commission in the Division of
9 Privacy and Identity Protection. With me is my
10 colleague, Katy Worthman, who is an attorney in our
11 Division of Financial Practices, and together we will
12 be co-moderating this panel.
13 Before I introduce our esteemed speakers, I
14 would like to remind everyone that we have staff in
15 the audience who have comment cards if you have
16 questions. We plan to make this interactive, make it
17 a conversation more than a presentation, and we will
18 welcome your questions and comments throughout the
19 panel and we will take them as they come.
20 So first, let me introduce our esteemed
21 panelists who have been so gracious to share their
22 time with us today. To my immediate left is Ryan
23 Calo, who is a Professor at the University of
24 Washington School of Law. To his left is Fred Cate,
25 Senior Policy Advisor for the Center for Information

1 Policy Leadership and a Professor at the Indiana
2 University School of Law. To his left is Jeremy
3 Gillula, who is the Tech Policy Director for the
4 Electronic Frontier Foundation, and to his left is
5 Irene Liu, General Counsel of Checkr. And at my -- at
6 the far end, last but certainly not least, is
7 Marianela Lopez-Galdos, who is the Director of
8 Competition and Regulatory Policy at the Computer and
9 Communications Industry Association. So welcome and
10 thank you.
11 So throughout the day, we have obviously
12 been talking about algorithms, artificial
13 intelligence, and predictive analytics. And the last
14 panel talked about ethical issues on those topics. In
15 this panel, we would like to drill down even more and
16 talk about the natural outgrowth of those ethical
17 issues, which are the consumer protection implications
18 for AI.
19 And with that, I would like to open up to
20 the panel to drill down into what are the consumer
21 protection implications?
22 MR. CALO: Should I start?
23 MS. GEORGE: Go ahead.
24 MR. CALO: Okay. Well, thank you very much.
25 I am honored to be here and really admiring of the

1 Federal Trade Commission's commitment to keep abreast
2 of emerging technology and a new leadership role in
3 that. One of the innovations of the FTC has been to
4 bring on technical staff very early so that they can
5 actually understand the technologies that they
6 regulate.
7 So I mean, you know, from a consumer
8 perspective standpoint, there are three I think
9 puzzles that I worry about. And they are each
10 about -- sort of about line drawing I guess you could
11 say. And the first is, does there come a point
12 whereby using machine learning and other techniques of
13 artificial intelligence that companies become -- have
14 such great information and power asymmetry over
15 consumers that we worry about advantage-taking.
16 So for example, the Federal Trade Commission
17 passed the door-to-door sales rule on the theory that
18 when someone comes to your house you are not in a
19 market context. I mean, this is a sort of much older
20 regulatory innovation, but the idea is that maybe you
21 are home and maybe you are in the middle of cleaning
22 or cooking or something like that and someone comes to
23 your house and tries to sell you something. Well, the
24 door-to-door sales rule is in recognition of the fact
25 that you are not in a consumer position right then.

1 So what it says is that you have abilities to unravel,
2 for example, the sale and you get certain other
3 things.
4 Well, what about the fact that increasingly
5 there are objects that are already in our house that
6 are doing the same thing? They are choosing when to
7 approach you. They are leveraging your hard-wired
8 responses to social interactions. Do we need a kind
9 of sales rule for, for example, the Echo? And so that
10 is one sort of -- I do not know how I should speak,
11 but, I mean, that is one example of where you sort of
12 worry about do we need special protections given the
13 intimate position that technologies increasingly have
14 within our worlds.
15 And then I have a couple of other puzzles,
16 which I will not get into such detail in because we
17 have a lot of people that want to talk, one of which
18 is, are standards of security sufficient? Because the
19 notion of security has been for a long time now the
20 idea that you are hacking into something and you are
21 bypassing a security protocol. But, today, lots of
22 machines can be tricked through a process called
23 adversarial machine learning, the idea being that
24 rather than bypass a security protocol, you just
25 purposely fool the system.

1 So to talk about Amazon again to keep with
2 the same example, researchers at Georgetown and
3 Berkeley showed that you could play some white noise
4 that none of us would think of as anything other than
5 white noise, but it would surreptitiously cause Amazon
6 to turn on the lights or to purchase something, and so
7 on. It was easily fooled in a way that was
8 problematic.
9 Our security standards, if you put something
10 out in the world that is easily tricked, a driverless
11 car that can be tricked into perceiving a stop sign as
12 a speed sign very easily, is that an unfairness
13 problem, much like having a system that is not secure?
14 Then the last thing, and I will stop here,
15 is I really worry quite a bit about the way in which
16 highly intimate information can be derived by what
17 feels to you like very ordinary information, the idea
18 that the intimate can be derived from the available.
19 It begins to break down this notion that somehow there
20 is sensitive information, personal information, and
21 that sharing it is problematic. You know, ultimately,
22 if things about you can you derived from what feels to
23 you like a mundane observation, because of the
24 extremely powerful tools of pattern recognition, you
25 know, perhaps we need to entirely rethink these

1 categories of sensitive and personal and so on.
2 So I will leave my provocations there for
3 now and pass it along, but thank you for the
4 opportunity to speak.
5 MR. CATE: Let me add my thanks. It is a
6 pleasure to be here and it is both important and
7 it is terrific that the FTC is doing this. I would
8 say I think we need to sort of start with some maybe
9 more basic principles, not about what the ethical
10 issues are, but rather about the ways in which we
11 raise them.
12 So one of those we need to recognize that AI
13 is already all around us being used in many ways. And
14 so a lot of today we have talked about AI as if it is
15 coming, as if it is the future, as if somehow we are
16 like ahead of the game in discussing fairness and
17 ethics and issues of consumer harm, whereas once again
18 this is a case where we are behind the curve as we
19 almost always are. It is almost impossible to be
20 ahead of technology. It is being used widely.
21 Second, I think as with many of the areas
22 involving information and certainly any time we talk
23 about privacy, we are already discovering that
24 people’s concerns are highly subjective and
25 contextual. So it really depends if we are talking

1 about my data or your data as to what my concerns are.
2 It depends on what the AI is being used for.
3 I wear an insulin pump. It uses a very
4 sophisticated AI to try to predict what is happening.
5 I hope it used all the personal data in the world and
6 continues to use all the personal data in the world,
7 but that is because it is keeping me alive. AI that
8 is being used to market to me, I might have very
9 different views about.
10 And then, third, I would just say I think we
11 will find in that same vein that the types of concern
12 that individuals have may be very different than what
13 society has. So what we know -- I mean, I think about
14 the number of people I know who work in privacy, who
15 spend their days talking about privacy, who really
16 care about privacy, who I know have a half-dozen or
17 more Echo devices at home. So individuals do not
18 always make rational choices and we should recognize
19 we might be concerned about something, but they are
20 voting with their feet and their pocketbook. They
21 know what they are getting into and they are doing it
22 anyway.
23 Finally, I would say there are the typical
24 set of concerns that we have with almost anything,
25 whether it is a refrigerator or a car, what have you.

1 Now, it, of course, involves data and that is that it
2 be reliable, that it be accurate to the extent it is
3 something that we care about accuracy. In other
4 words, I want the automatic brakes that use sensors
5 on my car to work consistently, I want them to work
6 only when there is something in front of me, and not
7 just to make it up and start slamming on brakes in
8 the middle of the interstate. And I want to have
9 recourse if they do not work. I want to know where
10 I can go, whether it is a court or the company or an
11 ombudsman or the FTC to get recourse when they do not
12 work.
13 MR. GILLULA: So we have heard some great
14 things already and I guess we will be jumping into
15 many of these things in more detail. So I will just
16 add two other things. So from a technologist's point
17 of view, I guess I think about two -- I have two other
18 things that I think I would add. One is from just a
19 process perspective in terms of doing consumer
20 protection, it is a lot harder I think to do consumer
21 protection when you do not have visibility into what
22 is going on.
23 So what I mean by that is AI offers the
24 ability to personalize things to a tremendous degree.
25 I mean, we have already seen this with targeted

1 advertising online. And it is very hard for an
2 outside organization like the FTC to see exactly what
3 ads people are being shown and based on what criteria,
4 unless the company that is actually showing those ads
5 makes a conscious effort, and some have. So to be
6 clear, this is something that is going on, but it is
7 an ongoing problem.
8 The other -- talking a little bit, Ryan
9 mentioned adversarial examples. The other thing --
10 and I think we will dive into this a little more, my
11 concern is just unintended errors, problems -- you
12 know, AI is great but it makes decisions in ways that
13 humans do not. So it can make decisions that no human
14 would ever make, you know, even without an adversarial
15 example and that no human would even be thinking that
16 an AI would make. So, you know, if that happens once,
17 you know, it is a one-off, it is an accident.
18 But then what happens when we are
19 replicating this across all society and we found out
20 that, you know, 1 percent of the time, it will make
21 some decision about a person. And if you talk about
22 the entire population of the U.S., now we are talking
23 about millions and millions of people who are getting
24 a very weird one-off decision. So I think we can talk
25 about that a little more, too. Thanks.

1 MS. LIU: So as a representative from the
2 industry, from our perspective, there is consumer
3 impact with AI regardless, positive and negative,
4 because there are mistakes that AI makes at times. So
5 the importance from our perspective as part of the
6 company perspective is that we need to make sure that
7 we analyze it up-front. So if you think about
8 privacy, back in the day there was a lot of discussion
9 around privacy by design and companies implementing
10 privacy by design, and how companies did that is they
11 implemented privacy impact assessments in a lot of
12 their products.
13 Similarly, it is very informative for
14 companies to implement AI by design. In a sense that
15 they should be assessing up-front because AI is out
16 there and we are using it in companies everywhere. So
17 understanding up-front with an impact assessment of
18 all of the different scenarios and how it can impact a
19 consumer in a biased way and in an unbiased way so
20 that you make sure that you understand up-front all of
21 the different scenarios and so that you can weigh the
22 probability and design it in such a way such that
23 fairness plays a role and that AI is not being used to
24 create mistakes or to make unfair decisions.
25 MS. LOPEZ-GALDOS: Sure. So please let me

1 take one minute to thank you for having me here and
2 also for putting together today's session, which has
3 been very informative.
4 As my initial remarks, I think one of the
5 things that we have learned throughout the day is that
6 AI is a catchall term. AI is going to be applied to
7 the credit score system, to the healthcare system, to
8 self-driving cars. So basically it is going to impact
9 all areas of society.
10 So when discussing and when drilling down
11 what ethical concerns we have and thinking about them
12 from a consumer protection perspective, I would
13 suggest to frame this discussion comparing machine
14 learning to the status quo. And what I mean by this
15 is that maybe we should try to talk about AI in the
16 context of healthcare and try to think whether there
17 is any difference to what we have right now and
18 whether the current regulations focusing on consumer
19 harm or privacy are sufficient to cover the same kind
20 of concerns we have, when machine learning is being
21 used.
22 And one of the things that we need to
23 acknowledge and -- sorry if I am being a little
24 pessimistic here -- but human beings and human
25 decisions are not perfect either. So we cannot hope

1 to have all decisions made by machines also to be
2 perfect. And some considerations that we might have
3 is that sometimes we might want to deploy AI systems
4 knowingly that they are imperfect because they bring
5 added value to humanity and balancing those tradeoffs
6 I think is going to be key for the future of machine
7 learning and deploying future technology.
8 MS. WORTHMAN: So in talking about the harms
9 that have come out of -- maybe more specifically in
10 the previous panels, people have spoken about bias,
11 they have spoken about privacy, they have spoken about
12 transparency. In looking at the current FTC
13 enforcement tools, FTC Act, Fair Credit Reporting Act,
14 the Equal Credit Opportunity Act, how well do these
15 statutes address the issues that have been raised by
16 these recent technologies?
17 And, Irene, I see you nodding, so I am going
18 to start with you.
19 MS. LIU: Sure. So Checkr -- for those that
20 do not know, Checkr is a background check company that
21 provides a platform to help companies hire faster and
22 in a more compliant fashion. So from our standpoint,
23 we are regulated already by the Fair Credit Reporting
24 Act. So when I think about regulations in AI and the
25 FTC Act in itself, I believe that the FTC Act is

1 drafted broad enough -- Section 5 is so broad in terms
2 of how it says unfair and deceptive practices. So it
3 is used in such a broad way that you could apply any
4 technology to it. So instead of developing
5 technology-specific laws, it is important for
6 regulators to keep in mind that companies like ours
7 and others have other regulations that are not just
8 FTC Act-specific.
9 So, for example, we have the FCRA Act that
10 requires us to comply with maximum possible accuracy
11 requirements, for example. So if we are producing a
12 report about you as an individual, we need to make
13 sure that it provides maximum possible accuracy. So
14 in doing that, when we are even implementing AI, we
15 need to make sure that AI technology is not making
16 mistakes, it is identifying the right person and that
17 it is creating the accurate report that we need.
18 So there are a number of other sectors like
19 ours that are governed by different laws. So if you
20 are in healthcare, obviously you have the healthcare
21 FDA laws, and if you are doing robo advisory from a
22 fintech perspective, there are SEC laws. So there are
23 a number of regulations that other companies are also
24 subject to that really put that checks and balances on
25 what companies can do with AI. So I think it is

1 important for regulators to think about that
2 holistically other than just the law that they are
3 regulating.
4 MR. CATE: So I think this is a great
5 question. I want to take the two laws you mentioned
6 separately. So the Federal Trade Commission Act in
7 Section 5, Unfairness and Deceptive Trade Practices --
8 actually, I have never met a regulator anywhere in the
9 world who would not like to have that authority
10 because of its breadth, because of the fact it is not
11 limited by a specific type of harm, because of the
12 reach, and, therefore, it applies to new technologies
13 without somebody having to update the law or say, "and
14 we mean artificial intelligence as well."
15 Now, having said that, it is kind of end-of-
16 the-road type of law. It does not tell you anything
17 up-front; it does not give you any prospective
18 guidance. These are things the FTC does in other ways
19 and other regulators do in other ways. So I doubt if
20 it is, if you will, going to be adequate to deal with
21 all the challenges that AI might present. But I think
22 it is a very broad flexible law, and in many ways, we
23 give it too little credit for its value in this area.
24 FCRA I actually think is discovering a new
25 birth, a new life. And again maybe not as exactly as

1 written, this may require some amendment, but this
2 notion of taking something where you use lots of data,
3 that data could be used in ways that affect people,
4 could be used in ways that would not affect people.
5 So you create some general obligations up-front, but
6 you make most of the significant rights, the real
7 actionable rights, depend on something happening,
8 something happening that would trigger an individual's
9 interest in saying, wait a minute, I may have been
10 disadvantaged or harmed -- and then other rights kick
11 in, you know, access to the data or a dispute, a
12 mechanism for dealing with accuracy, and so forth.
13 I think this could actually be a model that
14 we think of as we identify issues whether it is around
15 AI or big data or other types of intensive data uses,
16 a model for the future as well.
17 MS. LOPEZ-GALDOS: I think I am going to
18 tend to agree with what Fred and Irene just said.
19 From a European perspective, I think that the U.S. has
20 a technology-agnostic approach to consumer protection,
21 and I do not think that should change with AI because
22 of what I said in the beginning. It is going to
23 affect all aspects of our lives. And what I really
24 think we need to focus on is to see whether potential
25 consumer harms are covered or whether the laws are

1 sufficiently broad to tackle those, and if that
2 happens, then enforce the laws as they are. Some new
3 consumer harms might appear, but I believe that the
4 current system is sufficiently broad to cover those
5 probably. If not, I am sure you will find a way.
6 But I would not move towards a nontech-
7 agnostic approach. I think that could be bad for
8 innovation and that does not really make much sense if
9 what you are trying to resolve is potential consumer
10 harms. You should focus on whether consumers are
11 being harmed or not when thinking of regulations or
12 not.
13 MS. LIU: With that said, though, the FTC
14 could definitely play a role in providing guidelines,
15 not necessarily changing laws or creating laws, but
16 the FTC has been known to create guidelines in the
17 past, for example, security in the internet of things,
18 mobile security facial recognition, and those are some
19 of the aspects where the FTC did voice its opinion and
20 provided guidelines to businesses.
21 Especially in this area of AI where a lot of
22 companies are implementing AI and it is rapidly
23 moving, the FTC could influence in a way by providing
24 a guidance policy statement around their perspective
25 on AI and how to use it fairly and to create a fair

1 system that protects consumers.
2 MS. GEORGE: So following up on that,
3 obviously, the FTC in 2012 put out our privacy
4 framework and then a couple a years ago we did a
5 report on big data where we sort of laid out how
6 different legal laws that we apply, laws that we
7 enforce could apply in that area. Are there issues
8 that are unique to AI that are not covered by those
9 existing policy statements?
10 MR. GILLULA: No, go for it, go for it.
11 MR. CALO: I mean, I think it would be -- I
12 think we need to back up a little bit and say to
13 ourselves, okay, if artificial intelligence is as
14 powerful as proponents say, and if it is going to
15 remake society the way that proponents say, then also
16 we are going to need to have changes to law and legal
17 institutions. In other words, in my view, it is
18 either a bunch of hype or we are going to have to make
19 deep changes to our system. It cannot be like, oh, my
20 God, AI is going to change everything, but nothing
21 should change. That does not actually make a lot of
22 intuitive sense.
23 But let me just be more concrete. The kinds
24 of harms that I envision with artificial intelligence
25 that may be unique are twofold. There are wrong harms

1 and there are right harms. And the wrong harms are
2 when you get it wrong, and the line-drawing problem
3 that the FTC and others have to figure out is how
4 wrong do you have to get it, how easy is it to get it
5 wrong before there is a problem, you know. And that
6 is true whether it is wildly inaccurate, in which case
7 the credit reporting has something to say about it.
8 But I also think it is just like if something is
9 extremely easy to fool, even though in order to fool
10 the system you do not need to bypass any security
11 protocol, I wonder whether that might constitute
12 unfair design, in much the same way that designing
13 something that is really easy to hack might.
14 And then there are a set of right harms and
15 these are even harder. These are the kinds of harms
16 that happen when the technology actually is extremely
17 accurate. And we got to ask questions about that,
18 too, right? I mean, so what law, for example,
19 prohibits Uber from using Greyball to figure out
20 whether the people that are in the Uber are law
21 enforcement? You know what I mean? I do not know,
22 but that is an extremely innovative interesting new
23 thing to do is to use algorithms to figure out if
24 maybe the people in the car are going to be police and
25 then avoiding them, right?

1 And, yet, when the Federal Trade Commission
2 pursued Uber, you all pursued them along a very
3 similar lines to the way in which the FTC pursued
4 Amway decades ago. In other words, the big cardinal
5 sin originally for Uber was that it represented that
6 people were going to make more money on the weekend
7 than they actually were going to make, and that was
8 also Amway's big cardinal sin. But think about the
9 difference between Amway and Uber. I mean, these are
10 -- there is a sea change.
11 So I think that the Federal Trade Commission
12 Act is quite broad and unfairness and deception is a
13 dream at one level. It has some notice problems as
14 Fred alluded to. But what has to happen is
15 assertiveness. We need to make sure that the Federal
16 Trade Commission has the bandwidth and the mandate to
17 go in there and ask the hard questions, to direct
18 inquiries, and to figure out exactly what is going on.
19 Because I think one of the big problems is is that a
20 lot of the harms that are -- whether they are wrong
21 harms or right harms -- are invisible harms, and they
22 will not come to the fore unless the FTC uses its
23 authority to reach in and find out, or if, you know,
24 some reporter like Julia Angwin figures it out.
25 So, I mean, I do think we have adequate

1 tools and I think the FTC is precisely the right
2 agency to do it. But I think they need to be given
3 that mandate to, look, be assertive. This is a new
4 world. That is what we are being told. We are being
5 told this is a new world where everything changes.
6 Well, the FTC should change and it should pursue these
7 things very assertively. That is my own position. I
8 think you all are in the right agency to do it with
9 the right tools. But I think that that assertiveness
10 needs to come back.
11 MR. GILLULA: So the one thing I would add
12 to that is that transparency can help with that, too.
13 And it may mean that we need some sort of mandated
14 transparency when it comes to AI tools. Now, this is
15 not to say that we would want the same transparency
16 for all AI tools ever. It is going to be an entirely
17 different type of transparency for, you know, how does
18 your washing machine decide the optimal neural network
19 optimized way of washing your clothes versus, you
20 know, how does Uber decide whether or not you should
21 get a ride because it thinks you are law enforcement.
22 We definitely need some sort of content-
23 specific, but that could help an agency like the FTC
24 be able to see when the sorts of things that Ryan was
25 just talking about are taking place as if we had some

1 sort of mandated transparency.
2 MR. CATE: I think -- oh, go ahead.
3 MS. LIU: Go ahead.
4 MR. CATE: I think another way -- and just
5 to follow on Jeremy's point. You know, we have always
6 thought of transparency at least in kind of the data
7 or the data privacy world as meaning -- like
8 explaining what you are doing to people who frankly do
9 not care. So we have shoved notices down their
10 throat, we do not read them. We say, oh, we will make
11 them prettier, we will make them shorter, we will make
12 them layered. And at the end of the day, people just
13 do not read notices. That is just the reality. It is
14 a sad, but inconvenient truth.
15 So one thing we might think about is what
16 would transparency work like in this area. So part of
17 that might be documenting what you are doing. In
18 other words, it might be saying -- building a record
19 in exactly the way we require for human subject
20 research now. So, you know, we have the Belmont
21 principles that led to some law, if you take federal
22 dollars, you have to do this. You then have an IRB,
23 the Institutional Review Board, has to decide when you
24 are going to do things that affect humans. You have
25 to document it. You do not go to an agency to get

1 permission. I mean, the FTC would be overwhelmed if
2 that were the case.
3 But then if somebody bad happens, if humans
4 are injured, if something unexpected happens, then the
5 institution can be required to produce its
6 documentation that shows it followed a proper
7 procedure. It used the right calculation. Sometimes
8 bad things just happen even if you do everything
9 right.
10 So I think one of the things we collectively
11 need to be thinking more creatively about is what does
12 transparency look like in a field as rich and fast-
13 moving as AI and big data and other types of high
14 data-intensive fields and what it might be
15 supplemented with, so that we say, you know, maybe it
16 does not mean transparency to the end user who spent
17 all his or her life avoiding transparency, but rather
18 transparency so that it is available for a regulator
19 or for an advocacy group or if it is needed in
20 litigation or for other purposes.
21 MS. LIU: It is definitely important to
22 have that transparency. And so as companies are
23 building -- again, when I talk about that impact
24 assessment, it is important to think about audit-
25 ability and explainability not only to the consumers,

1 but also potential regulators. And I know Ryan
2 mentioned earlier that AI is huge and it is rapid
3 moving and so potentially the FTC needs a clear
4 authority on that.
5 From my perspective, if we start that route,
6 we are doing that with everything. I mean, everything
7 was big, mobile was big, internet of things is big.
8 So with every single new technology that emerges to
9 give FTC a clear authority on each one I think is
10 adding burdens and layers of enforcement -- the broad
11 enforcement that they need and that they already have.
12 So from my perspective, while it is
13 important to have that transparency, enforceability,
14 audit-ability in the companies for AI, in general, I
15 just do not think that we should be creating
16 technology-specific laws or enforcement mechanisms
17 within the FTC for specific technologies because there
18 will be new things that will be rapidly emerging again
19 and we will say this is the next big things. So at
20 that point, do we build another framework then?
21 MS. LOPEZ-GALDOS: I was going to react to
22 the discussion taking place right now and say three
23 things. First, I am a big fan of the FTC, so of
24 course they should have the mandate. I think that is
25 the case when consumers are being harmed. And that is

1 respective of whether the harm to consumers is being
2 produced by machine-learning technology or not. I am
3 going to support the technology-agnostic approach to
4 it to be able -- we protect consumers, which is what
5 we care about here.
6 Then with respect to the tradeoff between
7 accuracy and explainability, which I think is a very,
8 very hard balance to make and a hard analysis to make,
9 I think this is not new. Think about, for example,
10 gender-based price discrimination when it comes to
11 paying for car insurances. Well, people tend to
12 pay -- women tend to pay less than men because
13 basically it is easy to predict based on gender who is
14 going to have more accidents or not. So not
15 everything is new. Some of those tradeoffs and some
16 of the hard analysis we need to make between accuracy
17 of systems and explainability, we are already thinking
18 about them and they already exist in our society.
19 And the last point I wanted to make is that
20 with respect to transparency, I think it is important,
21 very important, because these systems are very
22 complicated, but I also think we need to have an
23 approach where the different degrees of transparency
24 exist. So for example, if I go to the doctor and what
25 I am trying to find is whether I have breast cancer or

1 not, I do not think I need to know how the machine
2 created all the neural networks to find out that I am
3 going to have breast cancer. I just want to know it
4 is accurate or not and just have a treatment, whereas
5 the doctor might need a different degree of
6 transparency to be able to ascertain the diagnosis.
7 So I think we need to bring the transparency
8 debate to a more down-to-earth or a more reality-based
9 approach and analyze it on a case-by-case basis.
10 MR. CALO: I guess -- I mean, first of all,
11 I am not arguing -- personally, I am not arguing that
12 the Federal Trade Commission should get AI authority.
13 It would be kind of cool, you could get little badges
14 with AI division.
15 (Laughter.)
16 MR. CALO: That is not what I am arguing.
17 MR. GILLULA: Would they say “robocop” on
18 them?
19 MR. CALO: They would say “robocop” on them.
20 This is ingenious.
21 I mean, I think that what I am saying rather
22 has to do with just how inquisitive the agency is,
23 right? So imagine that we are talking about -- you
24 know, not talking about consumer harms for a moment.
25 We are back now in -- we are talking about people

1 making crystal meth in their houses, you know what I
2 mean? And imagine the way that we regulated that
3 would be we say, listen, take a list of the
4 ingredients that you bought recently and post them in
5 front of your house, and if we walk over them and any
6 of them look like they might be the wrong ingredients,
7 then what we will do is we will follow up or something
8 like that.
9 No. I mean, there is a hugely different
10 stance when an agency -- a federal agency that has
11 been imbued with enforcement power, is asking pointed,
12 difficult questions, making you explain yourself.
13 There is a big difference between that and a kind of
14 transparency where you just sort of get to pick what
15 you want to share. You know what I mean?
16 Again, I do not think there should be a
17 special AI task force within the FTC exactly. But
18 rather I think that the FTC needs to use all of its
19 tools and I think that -- you know, listen, frankly,
20 just to speak plainly -- I have tenure now, so I can
21 speak plainly about things.
22 (Laughter.)
23 MR. CALO: You know, there has been a
24 history here where the FTC will pursue, more
25 assertively, consumer protection issues and then what

1 happens is Congress or the courts have placed limits
2 on that. So if I were the Federal Trade Commission, I
3 would be constantly thinking about what the right
4 balance is to strike, okay?
5 But we are in a moment. We have huge
6 companies calling for legislation, okay? We have
7 privacy legislation in California that we are going to
8 want to standardize, and so on. And so this is a big
9 moment, this is a time when we should be expanding is
10 what I am trying to say. But we have the tools and I
11 do not think we need to confer any special authority.
12 I just wanted to add that.
13 MS. GEORGE: So just a reminder to the
14 audience, if you have questions, please pass in a
15 card. This is a hot bench, so I am sure they would be
16 happy to answer whatever you wants to know.
17 I want to follow up a little bit on the
18 previous discussion. Ryan pointed out that we need to
19 go in and ask the hard questions in order to sort of
20 get to the heart of the matter. So I want to toss it
21 to Jeremy first as to what are the hard questions that
22 we need to answer in order to increase transparency
23 and explainability.
24 MR. GILLULA: So I was actually -- just as
25 you said that, I was thinking I was going to answer an

1 entirely different question. That is okay. In terms
2 of answering, you know, what are the hard questions
3 about explainability and transparency, I think I agree
4 quite a bit with Fred about -- that transparency to
5 the end user probably is not the right solution. We
6 have seen lots and lots of that and we have seen lots
7 and lots of it fail.
8 I am actually going to just use my
9 prerogative and answer a slightly different question,
10 which is what are the hard questions that the FTC
11 should be asking not just about explainability and
12 transparency, but about bias and fairness because that
13 is one that I have been thinking about a lot lately.
14 And I think the right answer there is, if
15 you are talking about a product or a service that has
16 a material impact on someone's life -- and I am going
17 to use that definition pretty broadly; I am even going
18 to include online advertising in that sense -- I think
19 the question you should be asking is what sort of de-
20 biasing or what sort of fairness calibration, what
21 sort of technical measure did you use? What
22 definition of fairness are you using?
23 Not, you know, we are going to say you must
24 use demographic parity or equality of opportunity or,
25 you know, any of these types -- but we are going to

1 ask which one you picked and did you do the
2 appropriate calibration because if you are not
3 thinking about how you can de-bias the results of your
4 algorithm in some way, then you are really not -- you
5 are clearly not thinking about the problem hard
6 enough. So I would throw that one out there as that
7 is the tough question that the FTC should be asking.
8 MS. WORTHMAN: Following up a little bit on
9 that, though, is there a risk that the black box of AI
10 is so complicated that you cannot identify what is
11 causing any of the bias?
12 MR. GILLULA: So it is --
13 MS. WORTHMAN: Or how to correct it.
14 MR. GILLULA: So the neat thing about the
15 correction part is there is actually a lot of active
16 research or rather in the last couple of years, some
17 papers published about how to take any black box
18 algorithm and correct it to some level. You know you
19 pick some certain type of fairness metric -- and to be
20 clear, by this, I am talking about a mathematical
21 fairness metric that says we want the same rate of
22 false positives or we want the same rate of false
23 negatives.
24 As we heard earlier today, there are many,
25 many of these. I think at last count I saw some paper

1 that said there was like two dozen different ones you
2 could choose from. Many of them are incompatible with
3 each other. But you can pick one and you could do it
4 post hoc. You do not need to actually go in and tweak
5 the algorithm. You can do it after the fact to the
6 algorithm.
7 So I am not too worried about the black box
8 nature or the explainability part of AI. I mean, that
9 was another thing we saw earlier today, too, was -- I
10 think it was a gentleman from Google who was showing
11 how they had done some really neat research on
12 explainability for AI systems, including
13 visualization. So I really do not see the -- for
14 me, the lack of explainability about AI is that
15 companies generally do not want to share information
16 about their algorithms because they are worried that
17 they will lose their secret sauce, and I totally
18 understand that. But it is not about that the
19 algorithms themselves are somehow incomprehensible or
20 unexplainable just because they are on a computer.
21 MS. WORTHMAN: And this is a question from
22 the audience. Given the decentralized privacy
23 protections in the United States, how will consumers
24 be completely from harm from AI devices where the harm
25 falls outside the regulatory authority of the FTC?

1 MR. CATE: So I am glad you asked that.
2 First of all, consumers are never going to be
3 completely protected from harm and we should stop
4 talking as if it is possible. And that has always
5 been the case with individual decisions, as well. I
6 know we have had rampant discrimination in individual
7 decisions in credit, in policing, in admissions for
8 decades, for centuries. And so the notion that
9 somehow AI is going to eliminate all that and that is
10 the standard we should hold it to is just setting us
11 up for a fall. I mean, no one will ever -- we will
12 just get rid of AI and we will be the much poorer for
13 it.
14 Second of all, it is interesting that the
15 question couched this in terms of privacy, a word we
16 have actually not used much up here at all. When we
17 talked about possible harms, privacy was not a
18 prominent one. I mean, we have talked about lots of
19 harms that you might say relate to privacy. But it
20 was interesting, while Jeremy was talking, I think
21 thinking to do the things he is talking about, which
22 are really important, you need data, you need to keep
23 the data. The way you detect, for example, that you
24 are getting a biased result is because you have data
25 revealed the bias.

1 So we are going to have to recognize that
2 there are going to be some tradeoffs here. In other
3 words, we might say in order to deal with questions of
4 fairness or bias, we actually hang on to more data, or
5 to deal with accuracy, we actually have to hang on to
6 more data. So I think we should at least be honest
7 with each other about the amount of tension between
8 these various goals.
9 And then just the last thing I would say
10 is the question used the word "harm," which is a word
11 I have used a lot. I like it because nobody knows
12 what it means, so you can comfortably use it. Like
13 Ryan will go write a law review article about it by
14 tonight --
15 (Laughter.)
16 MR. GILLULA: -- and show why we are all
17 wrong. But the problem with harm is we do not really
18 know what they are. In other words, it is harm using
19 my data without consent. Is it used in a way that
20 causes me actual injury, physical injury, financial
21 injury, some sort of severe emotional injury? Is it
22 noncompliance with some law relating to data, is that
23 by itself harm? So one thing which I keep saying as
24 we talk about AI, we need to also be talking about
25 what are the things we are trying to maximize and the

1 things we are trying to minimize.
2 So what do we agree are benefits? That
3 conversation seems to be fairly easy. And what do we
4 agree are the bad things that we would like to
5 minimize? Because they are not going to be
6 consistent. So that is going to be controversial
7 conversation which, frankly, the FTC is in a good
8 position to help lead.
9 Oh, dear, here it comes.
10 MR. CALO: No, nothing is coming apart from
11 me decimating -- no, I am just kidding.
12 (Laughter.)
13 MR. CALO: No, nothing is coming. First of
14 all, I fundamentally agree with Fred that we have to
15 get over this idea that reflexively just because you
16 gather more information, that is bad. You know what I
17 mean? More information often is very good and it is
18 very good for consumers in many, many, many contexts.
19 I guess what I would say about harm, I mean,
20 take, for example, a relatively well-known phenomenon,
21 and I believe it was one of the test prep companies, I
22 think it was Princeton Review, was found to be
23 charging more based on zip code for test preparation
24 in Asian American communities, right? That feels like
25 the wrong thing to do and it feels like the kind of

1 thing where I would, if I had a magic wand, go and ask
2 a lot of pointed questions about what other players
3 are doing in this space, what other metrics are using
4 to charge differential prices, and so on. Right?
5 And the harm, of course, is that because you
6 live in a particular neighborhood and there are
7 certain assumptions about the way that you value test
8 prep, you are paying more money. Sometimes it is not
9 at all hard to see the harm. The harm is just you are
10 paying more money, or with the lifetime value score
11 that The Wall Street Journal and later NPR discussed,
12 the idea that you might be on hold for a very long
13 time because you have a low LVS. These items are
14 pretty tangible. They are not well understood, and I
15 want us to be knocking on that door asking lots of
16 questions about these kinds of practices.
17 MS. LIU: At the same time, AI is not driven
18 by just PII. So there is a lot of data that we are
19 collecting that is anonymized, that is aggregated.
20 And so from a privacy perspective, it may not raise
21 privacy concerns. So it is really important to
22 differentiate those that are creating the -- using
23 information that may not be personally identifiable to
24 you to better your life. And so in that sense, I
25 think it is important for, as we are looking at

1 enforcement mechanisms, to think about privacy,
2 whether it is really impacting the individual, the
3 consumer.
4 And, secondly, again, you know, I talk about
5 AI by design and also an AI impact framework. And so
6 in that same sense, I really love Google's principles
7 around AI. One of the things that they also emphasize
8 is the importance of privacy by design when you are
9 developing AI frameworks. So that is something that
10 companies should do and I think this is a policy
11 guideline that potentially the FTC can encourage
12 companies to use, just like how it has done before in
13 terms of encouraging privacy by design in AI
14 frameworks as well, too.
15 MS. LOPEZ-GALDOS: Yeah, so I tend to agree
16 with that, but also when we discuss privacy, we need
17 to understand that privacy means a lot of things to a
18 lot of people and the value of privacy changes on a
19 consumer-by-consumer basis. Like if you ask people
20 whether they care about the environment, on climate
21 change, probably everybody -- almost everybody these
22 days will say, yes, I do care, but then not everybody
23 recycles. So we also need to understand when
24 consumers act rationally or not to discuss the privacy
25 requirement and what degree of privacy we want to

1 protect.
2 Because I am thinking of -- going back to my
3 previous example, I think everybody would like to be
4 able to use AI to identify potential cancer and to be
5 able to have a more accurate approach that determine
6 whether you are going to be sick or not well in
7 advance, as we saw examples earlier today. I am so
8 sure that people do not want to have their medical
9 records disclosed. And I think that tension is what
10 we need to look into and try to see whether the
11 current laws allow us to ensure that the consumers
12 have their, for example, medical records preserved,
13 which I think we can with the current laws and whether
14 -- how to make sure that society takes advantage of
15 AI, for example, advance the technologies that help us
16 identify potential cancers for all of us.
17 And I think the discussions need to be, as I
18 said in the beginning, brought to real cases and have
19 honest conversations about what we want and what we do
20 not want because AI can bring a lot of advantages for
21 society and we do not want to stop those. We
22 certainly do want to protect certain privacy elements,
23 for example, medical records, et cetera, but we need
24 to do it on a case-by-case basis and make sure we do
25 not impair the incentives to progress with these

1 technologies for the good of everyone.
2 MS. GEORGE: So it is interesting that you
3 talk about medical records because I think it was last
4 week I was watching our big data hearing and someone
5 said like most health-related information that is
6 available is not necessarily protected information, it
7 is more commercially available information. And so I
8 am just wondering if AI can apply in that sort of
9 space or how would you design protections around AI in
10 a space where many levels of information are not
11 protected in the traditional sense or where you can
12 infer data from someone from a nonprotected data set?
13 MR. CATE: So I would argue -- this may be
14 answering a different question -- but that it is not
15 very valuable to be looking at the data; it is much
16 more valuable to be using at the use and its impact on
17 the individual. So it does not matter whether I get
18 your health record or whether I figure it out from the
19 way you use your iPhone, if out of that, I make a
20 conclusion about your health status and I do something
21 with regard to that, presumably the impact on the
22 individual, for example, if it affects insurance rates
23 or it affects willingness of someone to employ you,
24 you know, uses that we would consider suspect, it
25 should not really matter the type of data, it should

1 matter the type of use.
2 I think AI is going to really drive that
3 home because we can make so many -- remember, AI is
4 all about probabilities and, you know, the probability
5 that that is your face, the probability that that is
6 the way it translates from, you know, Mandarin into
7 English, the probably that whatever, that you have
8 cancer, that you are pregnant, that you have some
9 other condition. And I think we are going to have to
10 stop worrying about where the data -- we may worry
11 about that for other reasons. Maybe there was a
12 promise the data would not be used or there is some
13 contractual issue that has to be dealt with. But
14 rather much more concerned about the use and the
15 impact on the individual.
16 MS. LIU: Companies should overall just be
17 thinking about what solution they are trying to drive
18 at with AI. So it is important at the design phase,
19 not only thinking about -- like I think a lot of
20 companies when they have data, they think about how
21 can I exploit this. And instead of using that
22 framework, it is important for companies to think
23 about what solution am I trying to solve, what use
24 case am I trying to solve. What user's life am I
25 trying to make better or easier? And what data can I

1 use from that to help develop a solution or a machine-
2 learning solution that can help better that life of
3 that user.
4 So with that context, they should also think
5 about what data do they need to collect, so collecting
6 only the data that is needed versus here is a data set
7 that I have, how can I exploit this. That is not
8 necessarily a right framework to go by from a company
9 standpoint, but rather thinking about solution-based,
10 and I think that will help drive solutions that
11 mitigate the consumer harm.
12 MR. GILLULA: I just want to completely
13 agree with Irene. From an engineering perspective, it
14 is also just bad statistics to say, I have the -- you
15 know, I found some data somewhere, now let me do
16 something with it because how you collect the data is
17 going to influence what data you have, which will
18 influence how accurate it is. And if you are going to
19 do something, if you say, well, you know, I want to
20 use it for some other purpose and so I will just --
21 you know, I know how to modify the records or I know
22 what portion of the data to throw out, then you
23 already sort of know what conclusion you are trying to
24 get.
25 I mean, I guess what I am getting at is, for

1 example, say I have some data set that I collected --
2 never mind. I was going to go into a pretty technical
3 example. If you are curious about that, I am happy to
4 talk with folks afterwards. Let me leave it at that.
5 (Laughter.)
6 MS. WORTHMAN: So, Jeremy, one of the things
7 that you mentioned previously was the fact that the
8 lack of -- like the availability of data actually
9 assists in identifying when there has been bias
10 implemented in AI. Could you discuss that just a
11 little bit in a particular instance?
12 MR. GILLULA: Yeah, so, I mean, so I think
13 -- so what I was talking about was that if you are --
14 if the purpose of the AI system is to do
15 personalization, so this is not here, now we are not
16 talking about systems that detect if there is breast
17 cancer or like the Adobe presentation that happened
18 earlier where I have some image and I want to find
19 similar stock photos. I am talking more about
20 targeted advertising or making loan decisions, that
21 sort of thing, where the only person who is going to
22 see, generally speaking, the result of some decision
23 is the person that decision applies to and whoever is
24 making the decision.
25 And so the concern here is that there is

1 just no visibility from the outside world. If I were
2 advertising 30 years ago and I chose to take out an ad
3 in certain magazines, then anyone can go pick up that
4 magazine and look and see what ads am I showing in
5 which magazines and am I showing certain ads to
6 magazines with certain demographics.
7 Now, it is a lot harder to do that. If I am
8 on Facebook or one of the other various online
9 advertising companies, it is much, much harder. And
10 then they are also doing all sorts of inference to
11 say, who is -- if I want to target people of a certain
12 demographic, with a certain background, with a certain
13 interest, some of that is going to be inferred data.
14 It is not actually going to be data that was actually
15 collected. And so it is even that much harder to be
16 able to tell, you know, am I doing something that is
17 having some unfair impact in some way?
18 MS. LOPEZ-GALDOS: Yes, I agree, but just a
19 clarification. There is users who decide when they go
20 and select online advertising who they want to target
21 and who they do not want to target. So it is not so
22 much the companies that do. So maybe the bias, we
23 find it in the user we want to target advertise. So
24 you have options. Do you want to target this zip
25 code? Do you want to target this audience? Do you

1 want to target -- there is like a list that you can
2 select. So I think when this cause bias, in that
3 respect, we also need to question ourselves when we
4 make the selections.
5 MR. GILLULA: Yeah, I mean, part of it does,
6 depending on the particular platform, fall on the
7 platform. So a good example of this is the current
8 complaints against Facebook that their housing and
9 employment ads, the framework was actually designed so
10 that it was easy to discriminate based on race. That
11 was a choice that Facebook made in how they designed
12 their platform and what characteristics they offered
13 in those sorts of advertising. It is totally true
14 that a lot of the time it is -- like is the person who
15 is buying the advertising, it is choices they are
16 making, but also some of it does apply to the
17 platforms and what choices they offer the person who
18 is buying the advertising.
19 MS. WORTHMAN: Building on that example, you
20 have the Fair Housing Act or you have the Equal Credit
21 Opportunity Act in the credit space where there is --
22 the FTC has enforced that law in the past. However,
23 taking those in the credit space or housing space sort
24 of out of that, when you have bias what -- this is a
25 question from the audience -- what general authority

1 does the FTC have to attack bias in the Section 5
2 context? Is it broad enough that it has been used to
3 attack bias on the unfairness authority?
4 MR. CALO: That is an excellent question. I
5 do not know who you are that asked that very good
6 question. No, I mean, it goes to the issue that Fred
7 and I were talking about, which is the idea of what
8 counts as harm, right? I mean, so especially under
9 the new -- newish, you know, decades old unfairness
10 standard, you have to weigh your regulatory
11 intervention against whether there is actual harm and
12 also you have to look at the benefit to society and to
13 consumers and the market.
14 So, for example, if you were to bring
15 something that you could show was societally valuable
16 and add a value to the market and to the consumer, but
17 also it had bias in it, even if we were to countenance
18 bias as being a harm, I do not think it would be so
19 obvious that that would constitute a problem, you
20 know. I mean, it is nontrivial.
21 What I will say is that I am a little
22 surprised that we are not talking a little bit more
23 about deception. In particular, I am a little
24 surprised we are not talking about the way in which a
25 lot of companies have way overclaimed about what this

1 stuff can do. You know what I mean? Way overclaimed.
2 So, I mean, for example, like I was in -- I am not
3 going to name company names, I was going to, but I am
4 not going to.
5 I was in an airport and I saw this
6 advertisement in the airport and it was just a bunch
7 of people that all looked similar to each other, like
8 it was like a cartoon. And then at the bottom it
9 said, artificial intelligence has already identified
10 who the terrorist is. No, it has not done that. That
11 is incorrect. It has not done that. That is a way
12 overclaim.
13 So sometimes people -- if you sell
14 nutritional supplements that do not do what they are
15 supposed to do or if you sell anything that does not
16 know what it -- usually you get in trouble for
17 deceiving. But for some reason we are giving these
18 folks that are advertising about AI a pass. I do not
19 understand why, right? I mean, there is verifiable BS
20 out there and I do not understand why it is not
21 deceptive.
22 MS. GEORGE: I have some more questions from
23 the audience. Can you describe new harms AI may
24 cause? And examples are synthetic video and audio and
25 virtual agents not identifying themselves as virtual.

1 MR. GILLULA: I can talk a little bit about
2 the virtual agents not identifying themselves as
3 virtual because Electronic Frontier Foundation
4 actually worked on a law in California that was
5 recently passed that was basically an online bot
6 labeling act. And the tricky part of this law, there
7 were bunch of problems with it, we got most of them
8 solved. One is what actually counts as a virtual
9 agent or what counts as a bot.
10 So let's restrict ourselves to social media,
11 say. Does it count as too much automation if I am
12 using something -- if I write a bunch of tweets and
13 then schedule them, is that too much automation and I
14 have to disclose that I scheduled them? What about if
15 instead I have a program, because I am a nerd and I
16 wrote up a program that will just automatically
17 generate tweets, but then I review each one and I pick
18 which ones to post? Do I have to disclose that -- do
19 I have to disclose that part? It gets into a very
20 hard line drawing exercise when you are talking about
21 what level of automation.
22 There are other parts, too, about if you
23 mandate things like an account has to disclose that it
24 is a bot. How do you enforce that? Basically, you
25 have to start unmasking people and then you get into

1 the harms of eliminating anonymous online speech. And
2 anonymous speech is something we value very highly in
3 this country. And if you are starting to eradicate
4 that online, you have to have a pretty good reason.
5 It looked like Fred was going to say
6 something so I am going to turn it over to him and we
7 will see where this conversation goes.
8 MR. CATE: I was just going to say I think
9 we are running the risk on this panel of being awfully
10 narrow in what we are thinking about as AI. In other
11 words, it is not just marketing and personalization
12 and targeted tweets. So AI is being used to deliver
13 healthcare. AI is the way we are examining MRI and CT
14 scan images. In other words, the harms we are talking
15 about are not -- a couple of weeks ago, I wrote a
16 letter to the president of a company because I
17 actually still believe presidents of companies love to
18 hear from me, and I got an answer back almost
19 immediately. I sent it electronically. And then I
20 spent the next like three weeks trying to figure out
21 was AI what did that, and I am absolutely convinced
22 that AI is what did it.
23 Was I harmed by the fact I got a nice
24 response that came from AI rather than the actual
25 president of the company who did not sit around

1 responding to my letter? This just, to me, does not
2 seem like the big issue. On the other hand, not
3 correctly diagnosing melanoma because we are using AI
4 to say is that image likely to be cancerous, that is a
5 harm. That is a really serious harm. Your car not
6 braking for a pedestrian, that is a serious harm that
7 is AI-related. We are using AI in some cities to
8 determine where police are based on calculations of
9 sophisticated data and realtime data about where
10 things are likely to go wrong. So not having police
11 where you actually need them, that is a real harm.
12 People will die because of that harm.
13 So as seriously as we can take the "I got
14 the wrong ad" or "I got a letter from the CEO that
15 really came from a virtual agent," I think we need to
16 be opening up our understanding of where AI is being
17 used in this economy, because it is massive. It is
18 being used to where we water crops and do not water
19 crops and it is being used to determine really
20 sophisticated life-changing things. I think it is
21 going to matter to the public frankly more than the
22 email they got. I am not criticizing the email, I
23 care about that. Can I send mine to you and will you
24 tell me did AI generate it?
25 MR. GILLULA: I can take a look, no

1 promises.
2 MR. CATE: Thank you.
3 MR. CALO: I just want to say that I wrote a
4 paper with a coauthor about the California disclosure
5 requirement that said that it has some First Amendment
6 issues with it. The truth is is that communicating
7 with bots is a new form of communication and one that
8 needs some breathing room. And I think that one
9 potential harm is that these emerging technologies
10 will freak us out and we will overreact. I think that
11 is personally what California did, and I think even
12 the current version, although because of the efforts
13 of the folks at EFF like Jeremy, is much, much better
14 than what it started out as.
15 I still think and I think my coauthor thinks
16 that it has some First Amendment issues. I mean, you
17 can go check out Regulating bot speech. It is coming
18 out in UCLA Law Review and see what you think. But I
19 think there are some real harms to overdoing it, too,
20 and I do not mean to be saying we should top down
21 regulate everything.
22 MS. LOPEZ-GALDOS: Yeah, we have seen some
23 examples of where some jurisdictions willing to
24 regulate up-front AI or the necessary elements for AI
25 to work and that is not necessarily, at this moment at

1 least, the right approach if we want to take advantage
2 of the full potential that machine learning has.
3 I think what we forget because now we are
4 hearing a lot about AI and machine learning is that,
5 yes, AI has existed for more than 50 years, but really
6 we are only in the nascent moment of the life cycle.
7 We are just beginning to understand the full potential
8 of it. If we start putting barriers to it, we might
9 not be able to allow the engineers to test and see
10 where this can take us.
11 So I think we need to, yes, worry about
12 consumer harm for sure. And FTC, you know, worry
13 about that and make sure that companies are able to
14 explain their systems and there is no bias, et cetera.
15 But this moment is really the beginning and let us see
16 where we go and let's have more workshops and let's
17 keep learning as we did today and see where the
18 technology stands. Today, this morning, we learned
19 that we talk about the full potential, but what
20 engineers can actually do at this moment is not the
21 full potential of AI. We are still working on the
22 systems and on deep learning, et cetera.
23 So I think it is very healthy to entertain
24 these discussions. It is extremely important to
25 probably do regular workshops on these matters. But

1 to cross the line and regulate everything, I think it
2 is just too early.
3 MS. GEORGE: This is open to the whole
4 panel. Are there particular contexts or uses where AI
5 should not be used since it is a nascent area? Should
6 there be a wait-and-see approach in certain instances?
7 MR. GILLULA: So it is not related to the
8 FTC's domain, but EFF, along with I think like 70 --
9 maybe 60 or 70 other civil rights organizations and
10 civil liberties organizations, signed a letter saying
11 that AI was not currently appropriate for bail,
12 parole, basically in the criminal justice context,
13 that we do not think the sufficient rules are there
14 and that -- and those are, as Fred was alluding to,
15 seriously life-impacting decisions. And so although
16 it is not in the realm of what the FTC would work on,
17 I think that is one that is important to note where
18 they are starting -- vendors are starting to market
19 and sell AI-related or AI-based risk assessment tools
20 and we definitely do not think it is appropriate.
21 MS. LOPEZ-GALDOS: I think I agree. As I
22 said earlier, I think the tradeoffs between
23 explainability and accuracy and that tension that
24 exists there is different whereas you apply AI to the
25 potential email you get or whether you are going to

1 incarcerate the person. So I think the debate needs
2 to be done on a very sector-by-sector basis and really
3 take accountability of the realities that that
4 decision is going to encounter. So, for example, if
5 as a result of applying AI, somebody is going to go to
6 jail and we cannot ensure that it is that accurate, I
7 would be more cautious than in other instances, for
8 example.
9 MS. WORTHMAN: Building off of that a bit,
10 depending on what type of AI is being implemented,
11 what choices and notice should consumers have
12 regarding the use of these types of technologies? I
13 mean, does it vary or should it be constant?
14 Irene?
15 MS. LIU: From a notice and consent
16 standpoint, it is definitely important. Most
17 companies also are regulated not only by the FTC Act
18 and others, but especially for those that are doing
19 business in Europe, there is GDPR as well. So there
20 is consent and notice requirement there especially
21 particularly with regards to use of AI. So it is
22 important to provide that notice to comply with GDPR
23 and it is also important to provide that notice for
24 transparency purposes from a consumer standpoint.
25 But what I liked about Marianela's

1 perspective earlier is how much transparency you want
2 to give to the users because it could be confusing.
3 So in the example that she provided earlier, the
4 doctor may want to understand what type of database
5 was used versus the patient. So in that context, you
6 do not want to flood consumers with too much
7 information about the type of AI and the database and
8 PII or even any type of information that is being
9 used. It needs to serve its purpose of providing
10 transparency, but not overtransparency that it causes
11 confusion and misleads consumers.
12 MS. GEORGE: So what does notice and consent
13 look like in an AI context? I will take an example I
14 think that many people can understand, which is credit
15 scoring and credit reports and it is built off of the
16 Fair Credit Reporting Act, which provides for access
17 to a copy of your report, dispute rights, things along
18 those lines. But in that space, you get a report. It
19 lists your credit lines and credit accounts or it
20 lists any criminal history you may have or your
21 educational history. It lists a series of items that
22 you can then look at and see whether or not they are
23 accurate and correct them if they are not. And once
24 those items are corrected, that will have an impact on
25 the ultimate decision. But in AI space how can you --

1 can you replicate that or what alternatives should
2 there be?
3 MR. CALO: Fred is a deep expert on notice
4 and choice, one of the leading experts on notice and
5 choice in America. But I will hazard something which
6 is that what is interesting about artificial
7 intelligence, at least when we come to embody it in an
8 agent, which is something that somebody asked about,
9 is that it can be awfully contextual and dynamic.
10 So I think that we ought to be encouraging
11 -- you know, the possibility of having a conversation
12 with Alexa about Amazon's privacy practices is, I
13 think, quite exciting, you know, and the idea -- maybe
14 you are anti, but the idea being that you could ask
15 specific questions rather than have some stupid thing
16 that was like really long and you are never going to
17 read it. But you could say, hey, Alexa, can Amazon
18 turn on you remotely to listen in on a conversation,
19 and then get an answer about that. I think that is
20 actually pretty powerful, personally.
21 MR. GILLULA: I am not anti, I just think
22 maybe only the people on this panel would find it
23 super exciting to have a conversation about Alexa
24 about --
25 (Laughter.)

1 MR. GILLULA: Which is not to say it would
2 be me, I agree. I just do not think the vast majority
3 of consumers would get a ton out of it.
4 MR. CALO: I mean, I think it is critical
5 when you are thinking about emerging technology
6 generally not just think focus on what is loss, but
7 what new affordances might be there or what you might
8 gain. I think that these things are quite powerful.
9 I think we are getting to a place where natural
10 conversations are becoming more viable and I think
11 that we should therefore -- I mean, if you think about
12 it, notice and choice, we have been operating under
13 basically Gutenberg technology all this time, right?
14 We just publish a long thing whether it is a
15 digital or a print, just a bunch of words on a page.
16 Yet, you know, here we have companies that are doing
17 these amazing things about organizing information and
18 gauging you and so on. Anyway, I think there is a lot
19 of innovation that could be occurring with notice.
20 And part of it would be to contextualize and actually
21 answer questions about this consumer instead of just
22 having something that no one reads.
23 MR. CATE: I would echo everything Ryan
24 said. I would just like to make two points. One is
25 we put in the record a paper that I did with some

1 colleagues at the Center for Information Policy
2 Leadership about AI, how it is used today and some of
3 the issues it raises, and one of the things we talk
4 about in there is the way AI is already being used to
5 enhance privacy protections, not just to make them
6 more easily understood or explainable, but to actually
7 activate them. So in other words, you can identify
8 somebody's privacy preferences as they start
9 expressing them and then you can start predicting what
10 they will be so that you offer them the default they
11 are more likely to care about. Rather than the
12 default that you want, you try to give them the
13 default that they want.
14 I would say just, in general, though, back
15 to the original question on notice and choice. As I
16 said earlier, we have relied on this largely because
17 we have not known what else to rely on for 50 years
18 now, with not a lot to show for it. And so I think we
19 should recognize that notice should be used and choice
20 only where there is something meaningful to tell the
21 individual and only where there is something they can
22 do about it. So I think it is terrific when my iPhone
23 says, did you know this app is using your contacts, do
24 you want to permit that? That is meaningful notice
25 and I can do something. I can say yes or no, I can

1 alter it.
2 But making my doctor add another paragraph
3 to the 65 paragraphs of the HIPAA notice saying, by
4 the way, your scans are going to be read by AI and, by
5 the way, you have no choice about that whatsoever
6 because it is far more accurate than humans, I am not
7 sure that is overly meaningful. I think we have to be
8 very contextual with notice because the effect when we
9 do not is that we just teach people to ignore all of
10 it. We get people in the habit of knowing that notice
11 is meaningless and so they do not read it, whereas if
12 we would use notice when there actually is something
13 worth telling them and something they can do about it,
14 we might resurrect notice as a meaningful data
15 protection tool.
16 Now, having said that, I am not disagreeing
17 with Irene. The law requires, both in Europe and in
18 some industries in the United States, notice and
19 choice, it is just bad law. In other words, it is
20 causing people to ignore these notices by providing
21 them when you cannot do anything about them and nobody
22 cares.
23 So one of the things we often talk about at
24 universities is, you know, a teachable moment. You
25 know, you can only teach someone when they are

1 interested in learning. Similarly, you can only give
2 meaningful notice when there is something that is
3 going to cause them to care about it. And that cannot
4 be they woke up in the morning or they went to a
5 doctor's office. It might very well be where there is
6 an event, there is something happening, there has been
7 some effect on them, there is some reason that they
8 would care, and then using the tools that Ryan was
9 talking about would be fabulous to really make notice
10 meaningful and interactive.
11 MS. LIU: There is always a conflict within
12 companies with product design when you are trying to
13 design products that is easy to use and that is easy
14 to understand. When you are throwing in all sorts of
15 consents and notices, it can make it really difficult.
16 And so there is often a conflict between the lawyers
17 and the product design teams about how can we make it
18 look beautiful without all your verbiage. So that is
19 something that we struggle with.
20 And I completely agree with Fred that
21 meaningful consent is ultimately more beneficial to
22 society and to consumers for how their information is
23 being used and how the company is using it versus just
24 providing our lengthy privacy policies that most
25 companies have.

1 MS. GEORGE: And as a corollary to that,
2 does the notion of opt-out work in an AI context and
3 does that vary based on I think the stage of the
4 product life cycle, be that data collection, you know,
5 product design when it is rolled out to market and
6 being used or other instances?
7 MS. LIU: Jeremy and I were talking about
8 this earlier. So from a GDPR standpoint, you do have
9 a right to erase your data. So there is an obligation
10 for companies to be able to remove that data. And
11 depending on how you configure that information, it
12 can be difficult. So that is something that you need
13 to think about from the beginning in the design phase
14 to ensure that companies, especially with the
15 California Privacy Act as well, it is important to
16 design these products in such a way that there is an
17 opt-out notion.
18 To opt out of AI, typically if a company --
19 if someone wants to opt out of AI completely, that is
20 like let’s say if I am using Netflix and I want to opt
21 out of using the choices, the different types of
22 videos or shows that they are showing to me, it is
23 basically opting out of using Netflix completely. So
24 you have to think about, like, are you trying to opt
25 out of the product or are you trying to opt out of the

1 database use as well? So there are different ways of
2 viewing opt out, and I think Jeremy can probably talk
3 more about the technological ways of opting out.
4 MR. GILLULA: Yeah, there has been some
5 recent papers that show that for neural networks you
6 can actually reconstruct what the training data was if
7 you are given sufficient time and access to -- and
8 able to run test data through the neural network,
9 which basically means that if I am a service and I
10 used your data to train my neural network, I cannot
11 remove your -- the fact that you are -- the imprint
12 your data has left on my neural network basically
13 without retraining it from scratch and retraining it,
14 once again, without your data. So it is technically
15 -- is it technically possible? Yes. Is that
16 potentially a huge burden on the company?
17 Potentially, yes.
18 Then there is the other question of, how
19 much benefit do you get from having your imprint
20 removed from whatever model was generated? Because it
21 does take quite a bit of effort to reconstruct all of
22 the training data, and so that is in the unlikely but
23 feasible attack. So there we do have to get sort of
24 into this balancing act a little bit.
25 MS. WORTHMAN: Another question from the

1 audience. In cases where autonomous systems result in
2 consumer harm, who should be held liable and to what
3 degree?
4 MR. GILLULA: Just send the robots to jail.
5 (Laughter.)
6 MR. CALO: Well, I mean, I think that is a
7 genuine puzzle. I mean, so you have -- in criminal
8 law and in tort law, we generally require that you do
9 something either on purpose or that you -- a
10 reasonable person would be able to foresee the
11 category of harm that occurred, right? And so when
12 you, for example, have a bot, which this really
13 happened, that is supposed to buy things randomly on
14 the web and buys methamphetamine and the police come
15 and say, you know, you bought methamphetamine, and you
16 say, no, no, it was the bot, right?
17 Or in another instance, where a company made
18 a bot that was arguably hacked into, but at least was
19 subverted by trolls that wound up denying the
20 holocaust which is not lawful in some jurisdictions
21 where -- that had access to this bot. You know, you
22 would be sort of hard pressed to bring a criminal case
23 to it. And certainly in many categories where --
24 something happens where the system just behaves in a
25 way that was not anticipated, you do not have what is

1 called proximate causation for purpose of bringing a
2 tort lawsuit, which is what I teach.
3 And that is not a great place to be because
4 you wind up in a situation where you have victims, but
5 not perpetrators. And I do not know how much that
6 would really matter to FTC enforcement, specifically,
7 because I think you could get around it just by
8 saying, look, you created these conditions that were
9 deceptive or unreasonable and these unexpected things
10 happened, but something was going to go wrong. But I
11 think it is pretty serious in tort and criminal law.
12 I think it is hard.
13 MS. WORTHMAN: We have also had a question
14 from the audience about retail price discrimination at
15 the individual consumer level and what is the material
16 harm to the consumer in price discrimination and maybe
17 price discrimination can be sort of whether or not it
18 is advertising different things, not on a prohibited
19 basis under ECOA, but just because you are using a
20 different type of computer, because you are purchasing
21 tickets on your mobile rather than on a laptop. What
22 is the harm, what is the cost-benefit analysis in that
23 particular instance?
24 MR. CATE: So this is a place where actually
25 notice would be quite useful. This would be much

1 more, in other words, to say if you visit this
2 website, we are going to use pricing based on
3 information about where you are coming from, the
4 computer you are using, whatever because it would then
5 empower you to say, well, I am going to go have my
6 friend check and see what the price is to see if I can
7 get a better price. In other words, that would be
8 actionable notice, you could really do it. And by the
9 way, having to disclose it would probably slow people
10 down -- companies down actually wanting to do that.
11 I mean, remember, we have discriminated on
12 price for forever, I mean, for generations. Every
13 time you fly, there could not be -- there is more
14 discrimination for all sorts of reasons, how long you
15 are willing to stay, what nights you will stay, and so
16 forth. We discriminate based on zip code, we
17 discriminate based on all sorts of other information
18 that have been imperfect. Now, we are going to be
19 able to discriminate better. I mean, we are going to
20 have both better technology and better data and the
21 two together are going to make much more precise
22 discrimination. You know exactly what I will pay.
23 eBay knows exactly what I will pay for something
24 because it has watched me pay that for years.
25 So this is actually a place where you could

1 say, first of all, we need to figure out is that a
2 harm. Is it something we are going to say is unfair?
3 Is it something that we are going to say causes
4 injury? And if not, maybe disclosure is sufficient.
5 To say, look, we are not willing to say we are going
6 to prohibit it, but we are going to say you get
7 notice. So now, you can figure out if you want to try
8 to come back at the system the other way. They are
9 doing it to you, can you do it to them?
10 But this is why we have to remember, again,
11 it is going to be very contextual and it is not
12 something new. It is not something AI is going to
13 create. AI is going to make it better in the sense of
14 potentially more precise or more tailored.
15 MR. CALO: I will give you my two favorite
16 examples of price discrimination after -- I mean, and
17 by favorite, I do not mean I like them. One of them
18 was a couple of years ago a marketing firm was using
19 this tool to figure out when women felt worse about
20 themselves and they labeled these “prime vulnerability
21 moments.” And they suggested that perhaps you should
22 advertise or charge people more during those moments,
23 you know what I mean? That strikes me as not a very
24 good use of price discrimination.
25 Another one of my favorites, although they

1 claim they never did this, is when Uber experimented
2 with figuring out whether you would be more willing to
3 pay surge prices when your battery was low on your
4 phone because maybe you would get stranded there.
5 Lovely, also. They say they have never done this and
6 I believe them about that.
7 The issue is not price discrimination. The
8 issue is taking advantage of people, which happens, it
9 happens a lot. And, yes, from an economic
10 perspective, better information is better. Maybe we
11 would worry at one level about all the social surplus
12 going to the firm. You know, they know your
13 reservation price. There is no windfall for you
14 because they charge you more if you would be willing
15 to pay more, so they get social surplus. We have
16 seemed to have moved away from the original
17 understanding of how consumer protection worked, which
18 was that it was immoral for firms to extricate all of
19 the social -- we seem to have moved away from that
20 model, and that is fine.
21 But I think it is the advantage-taking that
22 I really would worry about, and that is the kind of
23 thing I want there to be hard questions asked about.
24 MS. LOPEZ-GALDOS: Yeah. What I think is
25 that the questions we are addressing here, like from

1 the liability question and the answer from a total
2 perspective to this question right now is that there
3 are no new issues. Discrimination, price
4 discrimination has existed forever. It does not
5 matter whether a machine makes the decision or not,
6 the debate is the debate. We should analyze whether
7 we still -- whether price discrimination, for example,
8 is procompetitive or not or on the other side whether
9 consumers are being harmed or not, which approach we
10 want to take. But it is a debate that we should be
11 having and we have been having even without machines.
12 So I think we just need to continue talking
13 about these things, but I do not think it makes a
14 difference whether a machine makes a decision or a
15 human being makes that decision.
16 MR. GILLULA: So actually, I want to
17 disagree. There is something fundamentally different.
18 And if you lump in AI and big data and predictive
19 analytics altogether, then I agree there is nothing
20 new separate on AI. But a major difference is that
21 now there is a -- you are making a decision based on a
22 tremendous amount of data that has been collected as
23 opposed to just like, say, one data point that you
24 happen to notice or one data point you got like the
25 zip code or how many nights you want to stay for the

1 flight or something like that, something that is very
2 clear.
3 Now, you can potentially make price
4 discrimination decisions based on what websites the
5 person was visiting. Were they looking at budget
6 travel websites versus high-end travel websites? And
7 then there is the question of what happens if -- how
8 were the price discrimination decisions made if you do
9 not have any data on the person? And do they suffer a
10 penalty for preserving their privacy?
11 If I use a tracker blocker app on my phone
12 and I go to your website and I try to buy a plane
13 ticket and you do not have any history, am I
14 automatically categorized as I have to pay the highest
15 price or not as a punishment for not giving you data
16 about what level I might be willing to pay? So I
17 think that is a difference as opposed to say, you
18 know, what we have been doing for generations. It is
19 not different versus what we have been doing for the
20 last 10 or 15 years.
21 MS. WORTHMAN: Following up on that, though,
22 is there -- even though these are problems that we
23 have faced before, are there any particular harms that
24 are new based on price discrimination from AI or that
25 is a result of AI? Any new types of harms or is this

1 just the same thing that we have seen before?
2 MR. CALO: Well, I think there is a huge
3 difference. I think that -- again, I do think you
4 have to group together a bunch of different
5 phenomenon. It is not AI particularly. But, you
6 know, look, for a long, long time, companies have
7 noticed that -- and not just companies like mom and
8 pop shops, everybody has noticed, that there are just
9 cognitive limitations that we all have, right? We
10 just have these limitations to our rationality and
11 that is why everything costs $9.99, right? I mean,
12 obviously, okay?
13 There is a set of cognitive limitations that
14 behavioral law and economists, Ariely, Kahneman, and
15 so on have -- Christy Jolls at Yale -- have been
16 surfacing over a long period of time. And these are
17 things like prospect theory and status quo bias. And
18 sometimes the FTC actually intervenes and says, you
19 seem to be using status quo bias here with these
20 rebates. We are going to intervene because it does
21 not seem to be fair and you do not seem to understand
22 what is going on.
23 The issue is that even with all these
24 behavioral economists thinking about how we have
25 cognitive limitations, the list of cognitive

1 limitations is about 45 long, okay?. What artificial
2 intelligence permits you to do because it is so good
3 at pattern matching is to model what rational consumer
4 behavior would look like in a particular environment
5 and then look for deviations that are particular to
6 you, even if they are explicable. Turns out when you
7 are watching "Buffy the Vampire Slayer" on Tuesday
8 night, you are going to pay more for ice cream. I
9 know I am. But the point of the matter is that there
10 will be situations that are very, very specific to you
11 and perhaps not even have a theory behind them.
12 But what it allows is the mass production of
13 bias. That is what it allows. It allows these
14 systems to figure out where you are specifically
15 susceptible. And, indeed, we see early signs of this
16 already. I mean, you heard earlier a presentation
17 about how Netflix is showing different people posters
18 for shows based on guesses about their demographics or
19 their qualities. You know, that is part of the
20 phenomenon that in the literature is referred to as
21 persuasion profiling, the idea that not just that you
22 be matched to your interests, but that the messages to
23 you to sell you things would be matched to your unique
24 vulnerability.
25 So, for example, for some reason in your

1 life you are really worried about scarcity, well, that
2 advertisement will say, “while supplies last,” right?
3 And this is the kind of move that marketers are making
4 and it is only possible because of the way that we are
5 mediated by digital technology and we have these
6 intense analytic capabilities and, respectfully, I
7 think that is am enormous distinction from what has
8 come before.
9 MS. LOPEZ-GALDOS: So obviously, before, we
10 did not have self-driving cars and now, apparently, we
11 are going to have self-driving cars. So we are going
12 to see new things happening. Now, a self-driving car
13 might just cross over a person. What I was trying --
14 the point that I was trying to make is that the
15 thought process of analyzing the problems and
16 analyzing who is at fault, what was the causality, I
17 mean, the thought process is the same. The same that
18 exists without human beings is just applied to the new
19 setting.
20 I think the theories and thought process
21 should remain -- we should not think in the abstract.
22 We should think like we have a lot of analysis in tort
23 law, for example, and we want to say who is
24 responsible, who is not. In a self-driving car, there
25 is software, hardware, there are apps, there might be

1 somebody inside the car that was doing something as
2 well. And what I mean is that in the thought -- when
3 we are analyzing who is at fault and who is liable for
4 crossing over two people, the thought process of, for
5 example, causality should be the same as without AI.
6 That is an example -- for example, of the point that I
7 was trying to make.
8 MS. GEORGE: So I am going to ask one final
9 question and then I think we are going to wrap up. It
10 is going to be a compound question. Because I like
11 that.
12 So are there ways in which the FTC should
13 expand or rethink the notions of unfairness and
14 deception when it comes to AI and what educational
15 role should the FTC play with these new technologies,
16 both for consumers and businesses?
17 Marianela, do you want to start?
18 MS. LOPEZ-GALDOS: I think it is a very good
19 final question. I think the FTC is doing a great job
20 in putting together these hearings, as I said in the
21 beginning. I think AI is just a machine learning --
22 it is at a nascent moment. I think it is very
23 important to keep having a dialogue with businesses,
24 with the community, with the consumers, with experts,
25 and see where we are going to and see whether there is

1 anything that needs to be refined, for example, of
2 existing laws or not.
3 But what is very, very important is not to
4 think in the abstract of AI. We talk about AI as if
5 -- you know, at this moment, there are marvelous
6 things that can be done. I think there is a lot of
7 potential, but I really think that before stepping and
8 regulating or saying, oh, this is going to be a
9 disaster, everything is going to be mass-biased, et
10 cetera, we really need to understand where we stand,
11 what engineers can do, what companies are working on.
12 I think companies, at least the ones that
13 CCIA represents, are willing to cooperate with the
14 authorities, are willing to engage in adopting
15 principles. And I think having an open and frank
16 dialogue about what is going on is key to make sure we
17 get the right approach. So society can really profit
18 from AI.
19 MS. GEORGE: Irene, you just want to
20 continue down the line?
21 MS. LIU: Sure. Again from the beginning, I
22 feel that the FTC framework and the existing laws are
23 sufficient and the fact that it is broad enough that
24 it can capture AI, I think that is great. I think FTC
25 has withstood the test of time because it is broad.

1 But at the same time, I do think -- I agree with
2 Marianela that it is important for the FTC to continue
3 talking to the industry, also with other regulators
4 and academics to make sure that they are abreast of
5 this nascent technology.
6 There is also movements across the
7 globe, it is not just a U.S. phenomenon, but just
8 globally. Again, there is a recent universal
9 guideline for AI that was launched in 2018 by a
10 number of data protection officers recently. The
11 World Economic Forum is working on this issue.
12 Regulators in Europe, China, have taken a deep
13 interest in AI and so there is a lot of cross-country
14 developments within AI as well that the FTC can also
15 engage in to make sure that it stays ahead in terms of
16 the policy developments around the world so that we
17 are not hindering innovation, but fostering it as
18 well. So from that perspective, I think the FTC Act
19 is moving in the right direction with these types of
20 hearings as well.
21 From an education standpoint, the FTC can
22 also play a role in educating consumers to understand
23 what is AI. Again, because it is a new technology,
24 people hear about it. We talk about it all the time
25 in Silicon Valley, but it may not be known to the rest

1 of the country. So just educating people about what
2 chatbots are, what it means when you are choosing
3 Netflix on a Tuesday night and watching "Buffy the
4 Vampire Slayer," what the impact might be. It might
5 be that your ice cream prices might go up or it may be
6 that your Netflix fee might go up if you are a more
7 avid watcher than others.
8 So just understanding the impact of the data
9 would be helpful to consumers and also encouraging
10 companies to implement AI not just to exploit data,
11 but to think about it holistically is really important
12 and encouraging companies to do that from that
13 framework of advancing society versus exploiting the
14 data is something that FTC can take on, too.
15 MR. GILLULA: So I am actually going to
16 answer the question in reverse order. In terms of
17 consumer education, I think to accomplish that
18 mission, the FTC needs a much more robust staff of
19 technologists. They have only somewhat recently
20 started having technologists on staff. I feel like
21 the FTC should have as many technologists as lawyers
22 at this point. And, obviously, that is not where we
23 are.
24 I also realize that is not in the FTC's
25 ability to change. So if you are a Congressman or a

1 Congresswoman sitting in the audience, this is my plea
2 to you is increase the funding for technologists at
3 the FTC because those technologists can help with
4 explaining AI and what to expect in a consumer
5 standpoint to consumers. They can also help explain
6 it to the lawyers at the FTC when they are doing
7 enforcement actions or they are doing investigations.
8 They can help explain it to policymakers. So I think
9 there is a real need for a really robust technical
10 staff there.
11 In terms of whether or not the FTC Act
12 sufficiently captures everything that we might worry
13 about with regards to AI, I still worry a little bit
14 about the fact that -- I mean, I guess there are two
15 parts. One is whether -- I mean, at least -- and,
16 again, you got the only nonlawyer I think on the panel
17 talking. The FTC Act -- when you are talking about
18 harms and unfair and deceptive, you are talking about
19 what is the cost-benefit analysis. And I worry a
20 little bit that when we are talking about privacy, in
21 particular -- so, again, this comes back to rolling AI
22 and big data and predictive analytics into the same
23 thing.
24 But when you are talking about privacy, what
25 may be good for society is not necessarily good for

1 the most vulnerable part of the population because
2 privacy is really about privilege. You know, a
3 cis -- I am a cis, white guy, middle class, like I am
4 boring. Like you could know everything about me and
5 it does not matter because I am not worried about
6 something happening to me. But for many people with
7 very different demographics, they are very worried
8 about what data gets out about them.
9 And so while, on average, when we are making
10 that sort of cost-benefit analysis about what works
11 for society, that might make sense. But when we are
12 talking about privacy, we really need to be thinking
13 about what works for the most vulnerable part of the
14 population.
15 MR. CATE: So I think the FTC has enormous
16 capacity under Section 5 and FCRA and so forth. And
17 as Ryan was saying earlier, I think it should be
18 asking the hard questions and flexing those muscles.
19 Having said that, I actually do think additional legal
20 authority is likely to be necessary. Some of that may
21 be based more on what we might call procedure, but in
22 terms of ways that companies go about making decisions
23 and documenting those decisions about the use of
24 database automated decision making that affects
25 individuals in significant ways.

1 And then I do not think there is actually a
2 shortage of information, I think we have too
3 much information right now about AI and that one
4 role that the Commission might very productively
5 play, as it is doing now, is helping to sort of sort
6 through some of that information. I mean, everyone
7 on earth now has a code on AI. They all start with
8 fairness and have no idea what fairness means, not
9 the first idea.
10 And so helping to -- for example, as you
11 have begun today, thinking through what is fairness,
12 what are the elements of fairness, how do you measure
13 it, what is a desirable goal. The same thing about
14 harms. I do not think we have any agreement at all
15 about what are harms. I mean, we know the extreme of
16 harms. If someone is specifically injured or they
17 lose money, that is a harm. But what about between
18 where we are and there?
19 So in this area, I think the FTC has an
20 enormously important role to play and, frankly, a
21 great deal of experience to draw on in trying to kind
22 of sort through all of the stuff that is out there and
23 emerging and try to help make sense of it for
24 individuals and for businesses alike.
25 MR. CALO: Yeah, I mean, there has been a

1 lot of healthy back and forth and disagreement about
2 certain things on this panel, but I think that you are
3 seeing a rough consensus that the Federal Trade
4 Commission is well suited both because of its
5 expertise and because of its century of protecting
6 American consumers. I think we need an FTC that is
7 very assertive and uses the full range of its powers
8 and pushes the definition of unfairness and deception
9 and updates it for contemporary context. That is what
10 is so beautiful about a standard is that it can be
11 updated. And if these new technologies are as
12 powerful as people claim, so powerful that we need to
13 get out of their way, then they are also the kind of
14 thing that require a change to law and legal
15 institutions.
16 So my hope, too, is for -- I do not know
17 that there is any additional authority really needed.
18 I just think that the Federal Trade Commission should
19 be emboldened to pursue these very assertively and
20 that Congress and the courts should let them do their
21 job.
22 MS. GEORGE: With that, I want to thank our
23 panelists and audience for an exciting discussion.
24 I want to remind everyone to come back for
25 day two tomorrow for more interesting insights. And

1 thank you all for participating in this process.
2 Thank you.
3 (Applause.)
4 (Hearing adjourned.)
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

1 CERTIFICATE OF REPORTER
2
3 I, Linda Metcalf, do hereby certify that the
4 foregoing proceedings were digitally recorded by me
5 and reduced to typewriting under my supervision; that
6 I am neither counsel for, related to, nor employed by
7 any of the parties to the action in which these
8 proceedings were transcribed; that I am not a relative
9 or employee of any attorney or counsel employed by the
10 parties hereto, not financially or otherwise
11 interested in the outcome in the action.
12
13
14 s/Linda Metcalf
15 LINDA METCALF, CER
16 Court Reporter
17
18
19
20
21
22
23
24
25